APPLICATIONS AND CHALLENGES IN MASS SPECTROMETRY-BASED UNTARGETED METABOLOMICS

A Dissertation
Presented to
The Academic Faculty

by

Christina M. Jones

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Chemistry & Biochemistry

Georgia Institute of Technology
May 2015

COPYRIGHT © 2015 BY CHRISTINA M. JONES

APPLICATIONS AND CHALLENGES IN MASS SPECTROMETRY-BASED UNTARGETED METABOLOMICS

Approved by:

Dr. Facundo M. Fernández, Advisor
School of Chemistry & Biochemistry
Georgia Institute of Technology

Dr. John F. McDonald
School of Biology
Georgia Institute of Technology

Dr. Lawrence A. Bottomley
School of Chemistry & Biochemistry
Georgia Institute of Technology

Dr. Mark P. Styczynski
School of Chemical & Biomolecular Engineering
Georgia Institute of Technology

Dr. Julia Kubanek
School of Chemistry & Biochemistry
School of Biology
Georgia Institute of Technology

Date Approved: March 30, 2015

…life for none of us has been a crystal stair, but we must keep moving. We must keep going. And so, if you can't fly, run. If you can't run, walk. If you can't walk, crawl. But by all means, keep moving.

— Martin Luther King, Jr.

This is dedicated to
the women who always believed in me, inspired me to work hard no matter the circumstances, and taught me to enjoy life,
my mother Betty, grandmother Mary, and granny Lerlean.

Thank you for your unconditional love, support and faith.

ACKNOWLEDGEMENTS

Many people have been involved in the journey I began upon my matriculation at Georgia Tech. First I would like to express gratitude to my advisor, Dr. Facundo M. Fernández. You have been an excellent mentor and extremely supportive of my many endeavors. I knew the first time I met with you that you were THE advisor for me, and I could not be happier that the both of us choosing each other worked out so well. Thank you for your professional guidance and financial support in pursuing my dissertation research. Additionally, I would like to thank my committee members Dr. Julia Kubanek, Dr. Lawrence Bottomley, Dr. John F. McDonald, and Dr. Mark Styczynski. I am very appreciative of your advisement, the collaborations that were fostered, and your interest in my graduate education and research in addition to my career goals.

All teams are essentially dynamic, and the Fernández research group has been no different. Yet, despite changes, the morale and support of our group have remained outstandingly static. I would not have made it to this point without my fellow comrades. I would like to thank past group members Dr. Manshui Zhou, Dr. Asiri Galhena, Dr. Glenn Davis, Dana Hostetler, Deanna Snyder, Dr. Arti Navare, Dr. Mark Kwasnik, Dr. Leonard Nyadong, Adam Kaylor, Dr. María Eugenia Monge, Dr. Jennifer Pittman, Dr. Prabha Dwivedi, Dr. M. Julía Culzoni, and Dr. Chaminda Gamage. Manshui, thank you for teaching me when I first joined the group and passing the metabolomics torch to me. Glenn, your endless support, friendship, and still being available for words of encouragement means the most. Maru, I could write so much, but thank you—thank you for constantly pushing me to be a better scientist and for being one of the best friends I

have. And of course I would like to thank my current group members Rachel Bennett Stryffeler, Joel Keelor, José Pérez, Eric Parker, Xiaoling Zang, Scott Hogan, Dr. Martin Paine, Dr. Jay Forsythe, Dr. Matthew Bernier, Dr. David Gaul, Danning Huang, JasLynn Murphy, and my undergraduate student Laura Winalski. Rachel and Joel, you have been the best. It was just us for so long—I am going to miss the bond we share. Eric, take care of Xiaoling; keep her safe. El Capitan (Marty), you will always be my dumpling daddy. And Laura, you were the breath of fresh air that I needed to get to the end. Seeing science and research through your eyes brought me back to the joy and excitement I experienced when I first set foot in a research lab—one of the main reasons I pursued a doctoral degree. You are going to climb major heights and break down barriers, and I wish you nothing but the best.

My LSU family has continued to have my best interest at heart despite the hundreds of miles that separate us. I owe many thanks to my forever mentor Dr. Isiah M. Warner. Dr. Warner has exemplified the true meaning of mentorship—always providing honest advice and recommendations, absolute support, professional opportunities that have aided in my growth as a scientific researcher, and welcoming me into his family. Dr. Warner, you have inspired me to follow my passions, and I hope that I can fill your big shoes one day. The Office of Strategic Initiatives staff has been just as wonderful—Dr. Monica Sylvain, Lisa Swilley, Dr. Karin H. DeGravelles, Dr. Ginger (Powe) Redd, Dionne Sibley, and Melissa Crawford—in addition to Gretchen Schneider and Dr. Graca Vicente from the IMSD program. I was groomed well for graduate school success, and I am happy that I continue to be a source of encouragement and inspiration for the many students who followed suit after.

Finally, I would like to thank my family and loved ones— the people who really never forget to remind me of the magnitude of my accomplishments. My mother, Betty J. London, has always inspired and encouraged me to dream big and work hard. Your faith in my capabilities never cease to amaze me. My interest in research sparked because of the unfortunate tragedy that uprooted our lives, and I am glad that we now have something great to show for it. I would not be here without you. My grandmother, Mary L. Jones, embodies what it means to be a strong female. You carry the weight of the world on your shoulders, but maintain a spirit of perseverance. Our conversations are a source of motivation, and I am glad that I make you proud. My brother, Christopher "Jamaal" Eruchalu, is the male version of myself. I love and adore it. I feel as though I am getting the chance to see myself grow and mature and venture out into the world again. I am glad that I can always count on you no matter what. Know that the world is ours for the taking. Kristen, you have positively impacted all aspects of my life and I do not know where I would be without that. Every time I see my precious niece Natalie, I am reminded of the hard-working example I have to set and the mindset of passionately achieving goals and having compassion (among other things) that she has to see and experience. We are destined for greatness, and Natalie is as well. Sakina, hopefully you now know that I am not like Tommy, and I have a j-o-b. I love that you believe in me and just assume that I can do whatever I put my mind to; where would I be without that? I hope that when we reminisce on this phase of my life in the future, I have achieved an accomplishment of which my godchildren Jordan, Jayden, and Julia can be proud. Tiana, it has been nearly 2 decades—almost our whole lives. You have been supportive and available for every single thing with which I have had to deal, and I am more grateful

than you will ever know. We have to show Taran that anything is possible if he puts his mind to it. Chloe, we are all we have, and I have finally made it to the end. You are my never-ending source of encouragement and my pick me up when I am down. Meagan, Dr. Christina M. Jones has arrived. You claimed it before anyone else. You kept me excited about the outcome throughout this entire process—you are my rock. You, Chloe, and I have without a doubt made the best of our situations and are aiming for the clouds. Dr. Lauren A. Austin, *venimus, vidimus, vicimus*—we came, we saw, we conquered. I knew I had a partner when I met you during that visitation weekend. We have shared so many experiences, and you have reveled in all of my successes—for which I am eternally grateful. There is never a worry when you are present because of that take-charge attitude of yours and your unwavering support. I am following your trek and joining the ranks…we made it. And of course, I would also like to thank the rest of my family. You have all played a role in this achievement. We must celebrate and let the good times roll.

I am a person of faith, and I would be remiss if that was not acknowledged. God has blessed and bestowed many gifts upon me. None of this would be possible without Him.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

Da	Dalton
eV	Electronvolt
g	Relative centrifugal force
g	Gram(s)
h	Hour(s)
Hz	Hertz
i.d.	Inner diameter
in	Inch(es)
κ	Curvature
K	Kelvin
L	Liter(s)
m	Meter(s)
M	Molar concentration
min	Minute(s)
m/*z*	Mass-to-charge ratio
mol	Mole(s)
o.d.	Outer diameter
ppm	Parts per million
R	Radius
rpm	Revolutions per minute
R_t	Retention time
s	Second(s)
t_d	Drift time

v	Volume
V	Volt(s)
V_{pp}	Peak-to-peak voltage
W	Watt(s)
Δm	Mass error

LIST OF ABBREVIATIONS

AA	Arachidonic acid
APCI	Atmospheric pressure chemical ionization
API	Atmospheric pressure ionization
AUC	Area under the receiver operator characteristic curve
ANOVA	Analysis of variance
Avg	Average
BEH	Ethylene bridged hybrid
BPI	Base peak intensity
CA125	Cancer antigen 125
CAH	Congenital adrenal hyperplasia
CCD	Charge-coupled device
CE	Capillary electrophoresis
CE-MS	Capillary electrophoresis-mass spectrometry
CF	Cystic fibrosis
CFTR	Cystic fibrosis transmembrane conductance regulator
CI	Chemical ionization
CV	Coefficient of variation
CV	Cross-validated
DART	Direct analysis in real time
DART-MS	Direct analysis in real time-mass spectrometry
DAVID	Database for annotation, visualization and integrated discovery
DC	Direct current
DDA	Data-dependent acquisition

DESI	Desorption electrospray ionization
DESI-MS	Desorption electrospray ionization-mass spectrometry
DHA	Docosahexaenoic acid
DIMS	Direct infusion mass spectrometry
DGLA	Dihomo-γ-linolenic acid
DKO	Double knockout
DMS	Differential mobility spectrometry
EBC	Exhaled breath condensate
ECD	Electrochemical detection
EI	Electron impact
EOC	Epithelial ovarian cancer
ESI	Electrospray ionization
ET	Early-stage tumor
FDA	Food and Drug Administration
FDR	False discovery rate
Feat.	Feature
FWHM	Full width at half maximum
GC	Gas chromatography
GC x GC-MS	Two dimensional gas chromatography-mass spectrometry
GC-MS	Gas chromatography-mass spectrometry
GIST	Gas-ion separation tube
glog	Generalized logarithm
HDMS	High definition mass spectrometry
HGSC	High-grade serous carcinoma
HILIC	Hydrophilic interaction chromatography

HMDB	Human metabolome database
hPEBP4	Human phosphatidylethanolamine-binding protein 4
HPLC	High performance liquid chromatography
HRAM	High-resolution accurate mass
HSQC	Heteronuclear single quantum coherence spectroscopy
IACUC	Institutional Animal Care and Use Committee
IM	Ion mobility
IMS	Ion mobility spectrometry
iPLSDA	Interval partial least squares discriminant analysis
IR	Infrared
IRB	Institutional Review Board
IVDMIA	*In vitro* diagnostic multivariate index assay
I3C	Indole-3-carbinol
KEGG	Kyoto encyclopedia of genes and genomes
LC	Liquid chromatography
LC-MS	Liquid chromatography-mass spectrometry
LIPID MAPS	Lipid Metabolites and Pathways Strategy
LOOCV	Leave-one-out cross-validation
LPA	Lysophosphatidic acid
LT	Late-stage tumor
LV	Latent variable
LysoPC	Lysophosphatidylcholine
LysoPE	Lysophosphatidylethanolamine
MMCD	Madison-Qingdao metabolomics consortium database
MRM	Multiple reaction monitoring

MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
MSTFA	*N*-trimethylsilyl-*N*-methyltrifluoroacetamide
NMR	Nuclear magnetic resonance
No.	Number
OC	Ovarian cancer
oPLS-DA	Orthogonal projection to latent structures-discriminant analysis
PC	Principal component
PCA	Principal component analysis
PCa	Prostate cancer
PEG	Polyethylene glycol
PI	Photo-ionization
PLS-DA	Partial least squares discriminant analysis
PM	Probe mode
PM-DART	Probe mode-direct analysis in real time
PPARδ	Peroxisome proliferator-activated receptor-δ
PSA	Prostate-specific antigen
QC	Quality control
QTOF	Quadrupole-time of flight
Ref	Reference
RF	Radio frequency
RFE	Recursive feature elimination
RMSECV	Root-mean-square- error of cross-validation
ROS	Reactive oxygen species
S.E.M.	Standard error of the mean

SIMCA	Soft-independent method of class analogy
SRM	Selected reaction monitoring
SS	Stainless steel
SVMs	Support vector machines
TIC	Total ion current
TM	Transmission mode
TMCS	Trimethylchlorosilane
TMS	Trimethylsilane
TM-DART	Transmission mode-direct analysis in real time
TNF-α	Tumor necrosis factor-α
TOCSY	Total correlation spectroscopy
TOF	Time of flight
TWIMS	Traveling wave ion mobility spectrometry
T-wave	Traveling voltage wave
UPLC	Ultra performance liquid chromatography
UPLC-MS	Ultra performance liquid chromatography-mass spectrometry
UPLC-MS/MS	Ultra performance liquid chromatography-tandem mass spectrometry
UV	Ultraviolet
UV-vis	Ultraviolet-visible
VIP	Variable importance in projection
WNL	Within normal limits

SUMMARY

Metabolomics is the methodical scientific study of biochemical processes associated with the metabolome—which comprises the entire collection of metabolites in any biological entity. Metabolome changes occur as a result of modifications in the genome and proteome, and are, therefore, directly related to cellular phenotype. Thus, metabolomic analysis is capable of providing a snapshot of cellular physiology. Untargeted metabolomics is an impartial, all-inclusive approach for detecting as many metabolites as possible without *a priori* knowledge of their identity. Hence, it is a valuable exploratory tool capable of providing extensive chemical information for discovery and hypothesis-generation regarding biochemical processes. Untargeted metabolomic analysis is the first step toward designing targeted assays to study specific metabolic pathways, detect clinical disease, or monitor environmental phenomena. A history of metabolomics and advances in the field corresponding to improved analytical technologies are described in Chapter 1 of this dissertation. Additionally, Chapter 1 introduces the analytical workflows involved in untargeted metabolomics research to provide a foundation for Chapters 2 – 5.

Part I of this dissertation which encompasses Chapters 2 – 3 describes the utilization of mass spectrometry (MS)-based untargeted metabolomic analysis to acquire new insight into cancer detection. There is a knowledge deficit regarding the biochemical processes of the origin and proliferative molecular mechanisms of many types of cancer which has also led to a shortage of sensitive and specific biomarkers. Prostate cancer (PCa) is the 2nd leading cause of cancer-related mortality in men worldwide. Although the

introduction of prostate-specific antigen (PSA) screening has slightly decreased mortality, it is still a non-specific diagnostic serum marker as elevated concentrations are also indicative of benign prostatic hyperplasia and prostatitis. Chapter 2 describes the development of an *in vitro* diagnostic multivariate index assay (IVDMIA) for PCa prediction based on ultra performance liquid chromatography-mass spectrometry (UPLC-MS) metabolic profiling of blood serum samples from 64 PCa patients and 50 healthy individuals. A panel of 40 metabolic spectral features was found to be differential with 92.1% sensitivity, 94.3% specificity, and 93.0% accuracy. The performance of the IVDMIA was higher than the prevalent PSA test, thus, highlighting that a combination of multiple discriminant features yields higher predictive power for PCa detection than the univariate analysis of a single marker. Numerous discriminant metabolites were mapped in the steroid hormone biosynthesis pathway. The identification of fatty acids, amino acids, lysophospholipids, and bile acids provided further insights into the metabolic alterations associated with the disease.

Chapter 3 describes two approaches that were taken to investigate metabolic patterns for early detection of ovarian cancer (OC). OC is the 5th leading cause of cancer-related deaths for U.S. women. Due to the unavailability of reliable screening tests in clinical practice and the asymptomatic course through early stages of the disease, the majority of ovarian cancer cases are diagnosed as advanced, metastatic disease with poor survival. Early detection is thus crucial in reducing ovarian cancer mortality. First, *Dicer-Pten* double knockout (DKO) mice that phenocopy many of the features of metastatic high-grade serous carcinoma (HGSC) observed in women were studied. HGSC is the most common and deadliest subtype that results in 90% of OC deaths. Using UPLC-MS,

serum samples from 14 early-stage tumor DKO mice and 11 controls were analyzed in depth to screen for metabolic signatures capable of differentiating early-stage HGSC from controls. Iterative multivariate classification selected 18 metabolites that, when considered as a panel, yielded 100% accuracy, sensitivity, and specificity for classification. Altered metabolic pathways reflected in that panel included those of fatty acids, bile acids, glycerophospholipids, peptides, and some dietary phytochemicals. These alterations revealed impacts to cellular energy storage and membrane stability, as well as changes in defenses against oxidative stress, shedding new light on the metabolic alterations associated with early OC stages. In the second approach, serum metabolic phenotypes of an early-stage OC pilot patient cohort were characterized. Serum samples were collected from 24 early-stage OC patients and 40 healthy women, and subsequently analyzed using UPLC-MS. Multivariate statistical analysis employing support vector machine learning methods and recursive feature elimination selected a panel of metabolites that differentiated between age-matched samples with 100% cross-validated accuracy, sensitivity, and specificity. This small pilot study demonstrated that metabolic phenotypes may be useful for detecting early-stage OC and, thus, supports conducting larger, more comprehensive studies.

Many challenges exist in the field of untargeted metabolomics. Part II of this dissertation which encompasses Chapters 4 – 5 focuses on two specific challenges. While metabolomic data may be used to generate hypothesis concerning biological processes, determining causal relationships within metabolic networks with only metabolomic data is impractical. Proteins play major roles in these networks; therefore, pairing metabolomic information with that acquired from proteomics gives a more

comprehensive snapshot of perturbations to metabolic pathways. Chapter 4 describes the integration of MS- and NMR-based metabolomics with proteomics analyses to investigate the role of chemically mediated ecological interactions between *Karenia brevis* and two diatom competitors, *Asterionellopsis glacialis* and *Thalassiosira pseudonana*. *Karenia brevis* is a toxic dinoflagellate known to exude allelopathic compounds that directly inhibit the growth of species with whom it competes for resources. This integrated systems biology approach showed that *K. brevis* allelopathy distinctively perturbed the metabolisms of these two competitors. *A. glacialis* had a more robust metabolic response to *K. brevis* allelopathy which may be a result of its repeated exposure to *K. brevis* blooms in the Gulf of Mexico. However, *K. brevis* allelopathy disrupted energy metabolism and obstructed cellular protection mechanisms including altering cell membrane components, inhibiting osmoregulation, and increasing oxidative stress in *T. pseudonana.* This work represents the first instance of metabolites and proteins measured simultaneously to understand the effects of allelopathy or in fact any form of competition and highlights the ability of systems biology to shed light onto the nature of complex ecological interactions.

Chromatography is traditionally coupled to MS for untargeted metabolomics studies. While coupling chromatography to MS greatly enhances metabolome analysis due to the orthogonality of the techniques, the lengthy analysis times pose challenges for large metabolomics studies. Consequently, there is still a need for developing higher throughput MS approaches. A rapid metabolic fingerprinting method that utilizes a new transmission mode direct analysis in real time (TM-DART) ambient sampling technique is presented in Chapter 5. In this approach, the sample is deposited directly on a stainless

steel mesh that is held in the ionization region by a custom-built module. As a result, the DART plasma gas stream interacts with the sample in a flow-through fashion, which maximizes the sample-ionizing species interaction and minimizes variance in sample positioning. The optimization of TM-DART parameters directly affecting metabolite desorption and ionization, such as sample position and ionizing gas desorption temperature, was critical in achieving high sensitivity and detecting a broad mass range of metabolites. Ramping the ionizing gas desorption temperature further enhanced analysis by adding a simple separation dimension to this ambient approach. In terms of reproducibility, TM-DART compared favorably with traditional probe mode DART analysis, with coefficients of variation as low as 16%. The longer-lasting TM-DART signals enabled the acquisition of full scan and product ion accurate mass spectra in a single experiment, resulting in greater confidence in metabolite identification. TM-DART MS proved to be a powerful analytical technique for rapid metabolome analysis of human blood sera and was adapted for exhaled breath condensate (EBC) analysis. To determine the feasibility of utilizing TM-DART for metabolomics investigations, TM-DART was interfaced with traveling wave ion mobility spectrometry (TWIMS) time-of-flight (TOF) MS for the analysis of EBC samples from cystic fibrosis patients and healthy controls. TM-DART-TWIMS-TOF MS was able to successfully detect cystic fibrosis in this small sample cohort, thereby, demonstrating it can be employed for probing metabolome changes.

Finally, in Chapter 6, a perspective on the presented work is provided along with goals on which future studies may focus.

CHAPTER 1. UNTARGETED METABOLIC FINGERPRINTING: AN OVERVIEW

1.1 Abstract

Untargeted metabolomics is an impartial, all-inclusive approach for detecting as many metabolites as possible without *a priori* knowledge of their identity. This chapter presents an overview of untargeted mass spectrometry (MS)-based metabolomics which has become a valuable exploratory tool capable of providing extensive chemical information for discovery and hypothesis-generation studies regarding biochemical processes. Metabolomics is a recently emerging systems biology "omics" field that uses a holistic approach to comprehensively characterize the small molecule metabolites (< 1500 Da) in biological systems. A history of metabolomics and advances in the field corresponding to improved analytical technologies are described. The metabolome is estimated to be comprised of thousands to tens of thousands of chemically diverse metabolites at varying concentrations. These investigations can be challenging, and the main analytical platforms employed for contemporary untargeted metabolomics are reviewed; primary focus is placed on MS-based methods as they are utilized for analyses throughout the subsequent chapters. Untargeted metabolomics requires unique methodologies for sample preparation and mass spectrometric analysis in addition to chemometric data analysis tools; thus, a MS-based untargeted metabolomics workflow is introduced.

1.2 An Origin Tale of Metabolomics

Metabolomics is the methodical scientific study of biochemical processes associated with the metabolome which comprises the entire collection of metabolites in any biological entity. These metabolites are intermediates and end products of cellular metabolism. Thus, metabolomics is critical for probing changes in metabolite levels and collecting pertinent metabolic pathway information in an effort to investigate the biochemical fingerprints specific to particular cellular processes. The total number of human metabolites is currently unknown, but 4,000 – 20,000 are estimated.[1] Metabolome changes occur as a result of modifications in the genome and proteome, and are, therefore, directly related to cellular phenotype. Thus, metabolomic analysis is capable of providing a snapshot of cellular physiology (Figure 1.1). Metabolomic studies can be conducted in both targeted and untargeted manners. Targeted metabolomics detects and measures known metabolites or metabolite classes. Conversely, untargeted metabolomics—fingerprinting (often interchanged with profiling)—is an impartial, all-inclusive approach that detects as many metabolites as possible without *a priori* knowledge of their identity.[2, 3] Although fingerprinting and profiling are used interchangeably, profiling classically refers to the quantitative analysis of a particular class of metabolites or a set of metabolites involved in a specific biochemical pathway.[4] Untargeted metabolomic (or metabolic) fingerprinting is a valuable exploratory tool capable of providing extensive chemical information for discovery and hypothesis-generation studies concerning biochemical processes. Additionally, information acquired from metabolomics studies can be used to characterize disease states of organisms and monitor their responses to external stimuli.

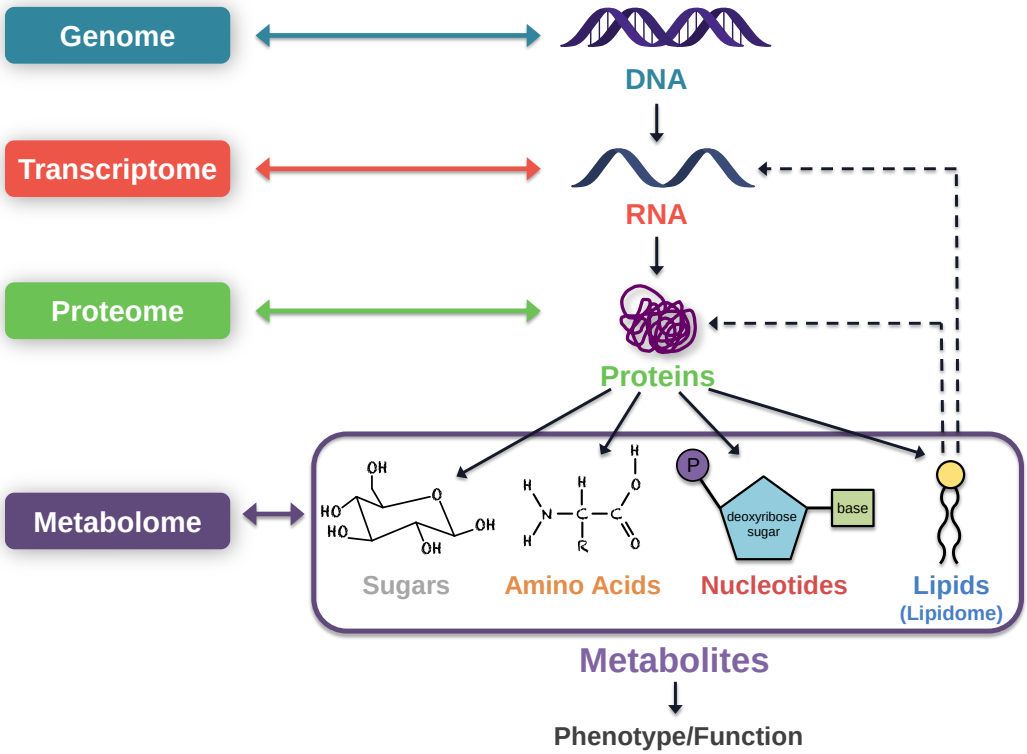

Figure 1.1: Omics Cascade. Metabolomics is the last level of the omics cascade and is more directly related to cellular phenotype.

Although metabolomics is still considered to be an emerging systems biology field, its origins can be traced back to traditional Chinese medicine. During 1500 – 2000 BC, traditional Chinese doctors employed ants to determine if urine glucose levels were too high in patients—an indicator of the metabolic disease diabetes mellitus.[5] These medical practitioners also detected diabetes by tasting urine for the sweetness associated with high glucose levels. While Chinese doctors did not have a name for this metabolic disease at that time, they recognized that characteristics of urine, a biological fluid containing many end and by-products of cellular metabolism, could be used to assess the state of health of any individual. During this same time period, Hindu practitioners of

Ayurvedic medicine also noted that sweet-tasting urine was associated with certain diseases.[5] Later in history (Middle Ages: 401 – 1500), ancient Greeks used 'urine charts' to link the tastes, smells, and colors of urine to medical illnesses.[6] Although contemporary metabolomics involves more sophisticated technologies, these ancient doctors correlated chemical patterns to biological processes—a fundamental notion upon which metabolomics was founded.

The growth of chromatography aided in the advent of contemporary metabolomics. This family of analytical separation techniques allow for simultaneous detection of many metabolites. C. E. Dialgliesh developed a two-dimensional paper chromatography method to profile urinary indolic compounds; indolic compounds may have been associated with bladder cancer and could provide information regarding vitamin nutrition.[7] From the late 1940s to early 1950s, Roger Williams and his research team utilized paper chromatography to acquire urinary and salivary metabolic fingerprints of many people.[8, 9] Subsequently, Williams found that similar groups of people, such as alcoholics or individuals suffering from mental disorders, may have characteristic metabolic fingerprints. Williams' work was a tremendous effort—more than 200,000 paper chromatograms were analyzed.[9] The availability of liquid chromatography (LC) and gas chromatography (GC) analytical separation techniques reduced the effort needed for such studies and, consequently, further expanded the field of metabolomics throughout the 1960s and 1970s.[10-14] Linus Pauling and collaborators believed that "information about the genetic nature of an individual human being, as reflected in the rates of various chemical reactions that take place in his body, usually catalyzed by enzymes, could be obtained by the thorough quantitative analysis of body

fluids".[14] Pauling developed a GC method to investigate biological variability by quantifying metabolites in urine and breath vapor based on the idea that "thorough quantitative analysis of body fluids might permit differential diagnosis of many diseases in a more effective way".[14] Additionally, Malcolm et al. demonstrated that metabolic profiles acquired from cerebrospinal fluid may be useful for the diagnosis of neurological disorders/infections in infants after profiles showed qualitative and quantitative differences between infant sufferers and controls.[12] Although GC and LC are very powerful separation techniques, neither has the ability to directly identify molecules in complex biological samples. LC was generally coupled to UV/UV-vis detectors and GC to flame ionization detectors, so analysis of chemical standards was required for metabolite identification. But, this changed when GC and LC techniques began to be coupled to mass spectrometry (MS) in the 1970s and 1980s. MS has undergone many technical advancements since the first modern-day mass spectrometers were created.[15, 16] These advancements greatly supported the expansion of contemporary metabolomics since numerous metabolites could be simultaneously detected and identified. E. C. and M. G. Horning developed GC-MS analysis methods to acquire urinary metabolic profiles of steroids, acids, and drug metabolites.[17, 18] In fact, the Hornings coined the term 'metabolic profile'. It was their belief that metabolic profiles "may prove to be useful for characterizing both normal and pathologic states, for studies of drug metabolism and the effects of drugs on human metabolism, and for human developmental studies".[17] The development and commercialization of atmospheric pressure ionization (API) techniques, chiefly electrospray ionization (ESI)[19] (for which Fenn shared the chemistry Nobel Prize in 2002) and atmospheric pressure chemical ionization (APCI),[20] allowed for the mass

spectrometric analysis of a broad range of analytes with varying chemical moieties and masses, in addition to the coupling of LC to mass spectrometers. As a result, the usage of LC-MS for metabolomics studies began to be on par with GC-MS.[21] In 1978, Henion presented research on the coupling of LC to MS to study drugs and their metabolites in biological matrices.[22-24] Games et al. utilized a newly developed interface which coupled LC to MS to metabolically profile plants.[25] In time, more and more clinical chemists began to realize the utility of chromatographic separation techniques coupled to MS for metabolic fingerprinting of biological samples and how impactful they could be for detection of inherited metabolic diseases and other diseases of interest.[12, 26] Additionally, clinical chemists and other researchers began to use multivariate pattern recognition methods to analyze and assist in the interpretation of their metabolomic data sets.[27-32] The importance of dimensionality reduction of the immense data sets and the need for sample class discrimination functions were recognized.

Concurrently with the development of GC-MS and LC-MS, advancements in nuclear magnetic resonance (NMR) spectroscopy instrumentation, first introduced in 1946,[33, 34] allowed for metabolite profiling of biological fluids.[35, 36] During the 1980s, ^{1}H-NMR analysis of urine proved to be successful in detecting several inherited metabolic disorders, including maple syrup urine disease,[37] methylmalonic academia,[37, 38] and phenylketonuria.[39] Nicholson et al. demonstrated that ^{1}H-NMR blood plasma fingerprinting could be used to monitor diabetes patients when they were no longer undergoing insulin therapy.[40] Similarly to hyphenated MS analysis techniques, contemporary metabolomics underwent tremendous growth once NMR spectroscopists

also began to use multivariate analysis techniques to interpret their substantial data sets.[6, 35]

From the 1990s to present time, metabolic fingerprinting technologies have incorporated MS- or NMR-based analytical platforms with multivariate statistical analysis techniques. New approaches have appeared for metabolic fingerprinting, such as capillary electrophoresis (CE)-MS,[41, 42] and improvements in chromatographic separation techniques, namely the development of ultra performance liquid chromatography (UPLC),[43, 44] have broadened the scope of metabolome coverage.[5] Although many technological advancements were made for comprehensive metabolite detection, a draft of the human metabolome was not made available to the scientific community until 2007.[45] David Wishart, along with numerous other scientists, catalogued more than 2,180 detectable endogenous metabolites using MS and NMR data collected from urine, blood, and cerebrospinal fluid, in addition to analyzing chemical standards and gathering metabolite information from books, journal articles, and electronic databases over a span of multiple years. The resultant database, the Human Metabolome Database (HMDB), has been updated since 2007, and now houses biological information for more than 40,000 metabolites. 'Detected' metabolites (those which have been experimentally confirmed) and 'expected' metabolites ("those for which biochemical pathways are known or human intake/exposure is frequent but the compound has yet to be detected in the body") are included.[46, 47] Moreover, the number of metabolites with biological fluid and tissue concentration data has greatly increased. The creation of HMDB has had a significant impact on metabolomics research— more than 1000 published scientific studies encompassing metabolomics, clinical biochemistry, and systems biology have

utilized the resources provided by the database.[46] Additional free databases used for metabolomics research include METLIN,[48] Kyoto Encyclopedia of Genes and Genomes (KEGG),[49] Lipid Metabolites and Pathways Strategy (LIPID MAPS) Structure Database,[50] and MetaCyc.[51]

1.3 Primary Analytical Platforms for Metabolic Fingerprinting

Comprehensively fingerprinting the metabolome is difficult due to its complex, diverse nature— thousands of metabolites belonging to different chemical compound classes, such as lipids, amino acids, and saccharides, exist. Moreover, these metabolites have dynamic levels ranging from picomoles to millimoles.[52] No singular analytical method can measure the thousands of metabolites estimated to be present in a biological system, as various analytical tools are successful in detecting different classes of metabolites. Although other analytical detection techniques are used for metabolomics research, such as ion mobility spectrometry,[53] and electrochemical detection (ECD),[54] MS and NMR (discussed in sections 1.3.1 and 1.3.2, respectively) are the primary technologies.

1.3.1 Mass Spectrometry

1.3.1.1 Gas Chromatography-Mass Spectrometry

The development of GC-MS was pivotal to the growth of metabolomics, and as such, is one of the most widely used analytical techniques for metabolomics studies. GC is typically coupled to MS via electron impact (EI) ionization. GC-MS is advantageous due to its high peak capacity, reproducible retention times, and ability to quantify metabolites.[4, 55, 56] Additionally, metabolite fragmentation patterns acquired by GC-MS

can be precisely replicated as EI ionization leads to very reproducible fragment ions, thereby, facilitating accurate metabolite identification by comparison to readily available compound databases, such as the National Institute of Standards and Technology mass spectral library. However, GC-MS is best suited for the analysis of volatile and thermally stable analytes. Polar metabolites must be chemically modified through derivatization reactions to reduce polarity and increase their volatility. Typically, acidic hydrogens in –COOH, -SH, -OH, and –NH functional groups are modified via alkylation, acylation, or silylation chemical derivatization reactions; additionally, carbonyl functional groups can be modified via methoximation derivatization reactions to inhibit enolization which can further complicate GC-MS analyses since multiple products would be present in resultant mass spectra.[4, 57] Methoximation derivatization followed by silylation of metabolites has been performed to broaden metabolome coverage.[58, 59] Nevertheless, chemical derivatization reactions can create undesirable artifacts and also have varying degrees of efficiency as sterically hindered analytes may only be partially derivatized.[57, 60] Moreover, added sample preparation steps decrease sample throughput for large metabolomics studies, thus, increasing analysis time.

Technological innovation has led to the development of two-dimensional GC-MS systems (GC x GC-MS). Typically, cryogenic modulation is utilized to transfer samples from a non-polar column onto a second polar column for rapid separation.[4, 61] In comparison to GC-MS, GC x GC-MS has substantial increases in peak capacity, chromatographic resolution, and sensitivity, thus, making it more suitable for metabolic profiling of complex biological samples.[61] Habram et al. demonstrated that GC x GC-MS analysis lowered detection limits of small hydrocarbons by a maximum factor of 27.[62]

For a metabolomics study on spleen tissue extracts, Welthagen et al. determined that GC x GC-MS detected 2.4 times the number of metabolites than GC-MS.[63] GC x GC-MS has been used to metabolically profile urine of infants to determine organic acids capable of diagnosing inborn errors of metabolism[64] and to explore metabolic differences in fermenting and respiring yeast cells.[65]

1.3.1.2 Liquid Chromatography-Mass Spectrometry

Although many of the first metabolomics studies utilized GC-MS, the growth and availability of high performance (HP) LC-MS aided in the expansion of this scientific field of study and is still one of the primary analysis methods used. While HPLC-MS has lower chromatographic resolution than GC-MS, a broad range of metabolite classes across a wide mass range can be detected with HPLC-MS by choosing optimal stationary and mobile phases.[56] Moreover, there is no need to derivatize metabolites prior to analysis. The separation of intermediate polar to nonpolar metabolites can be achieved by reversed phase LC-MS with C_{18} stationary phases being the most common. Hydrophilic interaction chromatography (HILIC) has been developed for the analysis of polar metabolites since they are not retained on reversed-phase LC-MS columns and elute with the solvent front.[66] Tolstikov et al. developed a HILIC-MS method that separated and detected highly polar metabolites in phloem exudates from petioles of *Cucurbita maxima* leaves.[67] They obtained the best separation of metabolites in the plant samples using an amide stationary phase. Oligosaccharides, glycosides, amino sugars, amino acids, and sugar nucleotides were detected. Reversed-phase ion-pairing LC-MS is typically used to detect charged metabolites;[68] however, unpaired ion-pair reagents can contaminate mass spectrometers and reduce the sensitivity and repeatability of LC-MS analyses.[55] Luo et al.

employed tributylammonium acetate as an ion pair modifier in reversed-phase LC-MS to identify intracellular metabolites involved in the central carbon metabolism (including glycolysis, the pentose phosphate pathway, and the tricarboxylic acid cycle) of *Escherichia coli*.[68] Sugar phosphates, nucleotides, and carboxylic acids were separated and detected. LC is primarily coupled to MS via ESI, a soft ionization technique that can be negatively impacted by matrix effects of complex biological samples.[69, 70] For metabolomics studies, positive and negative ESI modes are used to broaden metabolome coverage.[4]

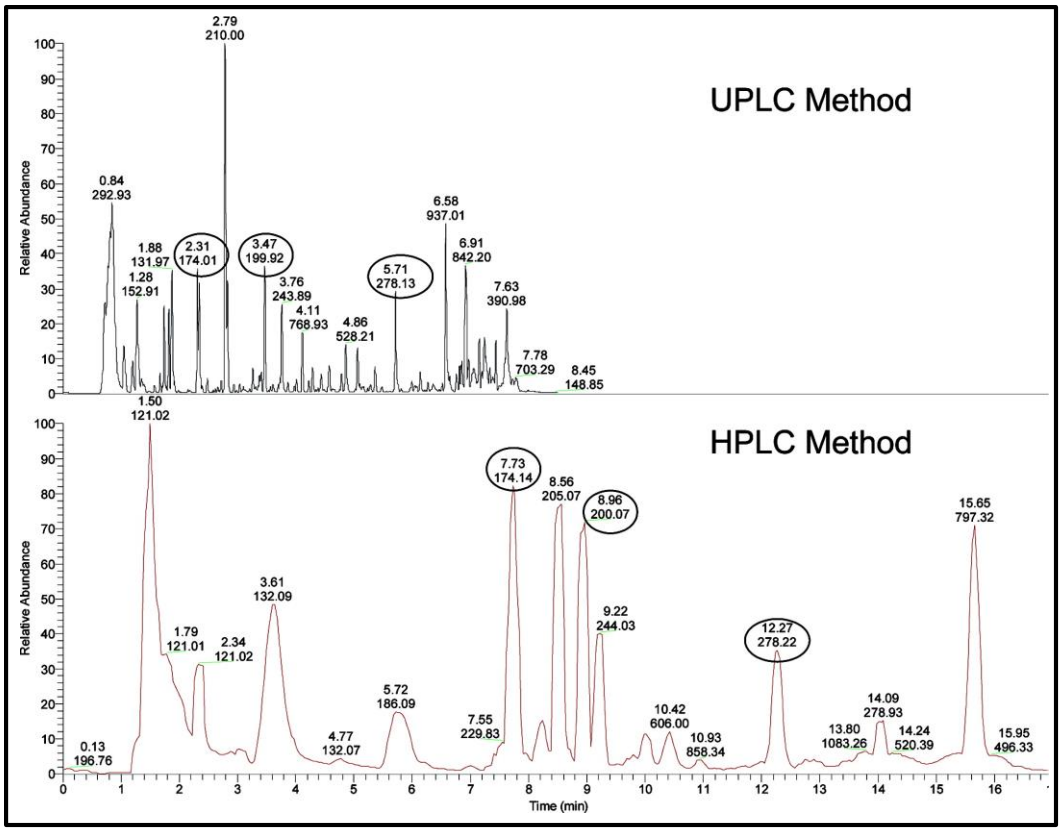

Figure 1.2: Comparison of base peak chromatograms from LC-MS untargeted plasma metabolite profiling. An increase in resolution was observed with UPLC-MS analysis along with a 3-fold reduction in analysis time. The three circled peaks represent common ions detected in both LC-MS methods for comparison of separation performance. Reprinted with permission from ref 73. Copyright 2009 American Chemical Society.

The introduction of UPLC-MS in 2004[43] greatly impacted LC-MS-based metabolomics.[55, 71] UPLC-MS utilizes 1.0 – 1.7 µm porous column particles, and, subsequently, requires instrumentation that can operate in the 6000 – 15000 psi pressure range.[43] Typical peak widths are 1 – 2 s and result in a 3 – 5 fold sensitivity increase when compared to HPLC-MS using 3 µm particle columns.[43] Moreover, increased peak capacity allows for higher chromatographic resolution as spectral overlap is significantly reduced. Rapid UPLC-MS analyses can be achieved without resolution losses which

increases the throughput of metabolomics studies.[43, 72] Evans et al. reduced LC-MS run time almost 3-fold when switching from HPLC-MS to UPLC-MS for the analysis of plasma metabolites, while still experiencing gains in chromatographic resolution (Figure 1.2).[73] This advantage allowed for two UPLC-MS data acquisitions (positive and negative ESI modes) in less time than a single HPLC-MS method with better precision. While analyzing the urine of Zucker rats, Wilson and coworkers detected ~1,500 ionic species during a 10 min HPLC-MS analysis, and over 5,000 ionic species in a 5 min UPLC-MS analysis.[74] A 1 min UPLC-MS analysis detected 1,000 ionic species, thereby, demonstrating that UPLC-MS can achieve similar results as HPLC-MS, but on a much shorter time scale. UPLC-MS has been used for metabolic profiling of human and animal tissues,[75] determining metabolic markers for hepatitis B detection[76] and hepatotoxicity.[77]

1.3.1.3 Capillary Electrophoresis-Mass Spectrometry

The first reported use of CE-MS for metabolomics research was in 2000.[41] CE separations are rapid and require little or no sample pretreatment.[78] Moreover, organic solvent consumption is extremely low or nonexistent and simple fused-silica capillaries are used instead of expensive LC columns.[78] CE is typically interfaced to MS via ESI[79-81] which is not a trivial feat owing to the low effluent flow rates from the capillary, the need to maintain electrical contact for electrophoretic current, and the usage of electrolytes in the buffer used during analyses.[82] CE-MS has been used to acquire metabolic fingerprints of human colon cancer cells,[83] and for the discovery of metabolic biomarkers for diabetic nephropathy.[84] Soga et al. employed CE-MS for serum metabolomics research that showed γ-glutamyl dipeptides may be potential biomarkers for liver diseases.[85] The γ-glutamyl dipeptides indicated reduced glutathione production. Although CE-MS is

capable of detecting a wide range of metabolites, it is best suited for the analysis of water soluble, charged metabolites. While reversed-phase ion-pairing LC-MS is capable of detecting charged metabolites, CE has less of a negative impact on the MS system.[55] Furthermore, CE has a higher theoretical separation efficiency than HPLC.[55, 56] CE is capable of analyzing metabolites in individual cells or very small sample volumes (nL up to a few μL).[78] Nemes et al. used CE-MS for metabolic profiling of six different neuron types from the *Aplysia californica* central nervous system.[86] They detected more than 300 distinct metabolites from a single neuron and were able to detect chemical similarities among some neuron types. The fused-silica capillaries used for CE-MS analysis can be coated to reduce electroosmotic flow, and, thus, enhance the separation windows for specific metabolite types. Ramautar et al. used coated polybrene-dextran sulfate-polybrene coated CE capillaries to ensure a larger separation window for cationic metabolites in urine would be achieved.[87] But, CE-MS can suffer from low repeatability when bare fused-silica capillaries are used as temperature changes cause deviations in migration times.[82] To circumvent this disadvantage, a genetic algorithm that aligns CE-MS data using accurate mass information has been designed.[88] Using murine urine samples analyzed by CE-MS, the algorithm showed a significant reduction in the migration time drift.

1.3.1.4 Direct Infusion and Ambient Ionization Mass Spectrometry

Direct infusion MS (DIMS) is a high-throughput MS analytical approach that involves infusing/injecting a sample directly into the ionization source of a mass spectrometer without employing chromatography. Since typical analysis times are no longer than a few minutes, DIMS is able to fingerprint a large set of samples rapidly.[89]

Furthermore, short analysis times ensure that instruments are stable over the duration of all experiments, and reduced technical variability improves the subsequent multivariate analysis.[90] Lin et al. compared DIMS and UPLC-MS for a kidney cancer serum biomarker discovery metabolomics study.[91] Twenty-three metabolites were found as potential biomarkers by DIMS analysis, while UPLC-MS analysis discovered 48. Still, DIMS had comparable multivariate classification and prediction capabilities as UPLC-MS, but only consumed ~5% of the analysis time. The authors demonstrated that DIMS has the potential to be a rapid diagnostic method, while UPLC-MS can be used limitedly when a comprehensive biomarker screening is needed. Mas et al. compared DIMS to GC-MS for metabolic profiling of yeast mutants and found that the data acquired from both methods was complementary.[92] GC-MS mainly detected amino acids and was well-suited for the classification of mutants with altered nitrogen regulation, while DIMS performed well at classifying mutants involved in the regulation of phospholipid metabolism. DIMS is used with API sources, however, ESI is the primary ionization technique used.[4, 89] Although DIMS is advantageous as a rapid analysis technique for metabolomics studies, it does have disadvantages. Isomeric species cannot be distinguished since they have the same mass, and there is a possibility for in-source fragmentation to occur as well as the formation of adducts.[4] Furthermore, DIMS suffers from ion suppression because all components of the sample are introduced to the ionization source at the same time.

Although DIMS is a rapid analysis method, sample preparation is still needed before metabolomics experiments are conducted to ensure the metabolites are in a liquid state. With the introduction of desorption electrospray ionization (DESI)[93] in late 2004 and direct analysis in real time (DART)[94] in early 2005, a new subfield of analytical

MS—ambient ionization MS—was established. Ambient ionization MS analysis is conducted in open-air and entails minimal-to-no sample preparation. Ambient ionization MS is used to analyze samples in solid, liquid, and gaseous states. The lack of sample pretreatment gives ambient ionization techniques an advantageous edge for metabolomics research since precious samples remain in their original states. Additionally, these techniques are high-throughput MS analyses, given that there is a lack of chromatographic separation and the analysis time is mainly influenced by the time needed to place the sample in the ionization region of the assembled instrumentation. Although ambient ionization MS is a recent subfield of MS, most ambient MS ionization methods rely on adaptations of ESI and APCI ionization mechanisms. Ambient MS has been utilized for metabolomics research. Cajka et al. determined the origin of selected beers using DART-MS metabolomic fingerprinting.[95] Pan et al. analyzed urinary metabolites by DESI-MS to identify patients with inborn errors of metabolism.[96]

1.3.2 Nuclear Magnetic Resonance Spectroscopy

The advancements in NMR analysis of biological fluids[35] aided in the growth of metabolomics; consequently, NMR remains one of the primary analytical platforms used for metabolomics research. NMR analysis is rapid, highly reproducible, and non-destructive, thereby, allowing samples to be analyzed in their crude form. Furthermore, NMR provides structural information for the detected metabolites. Unlike MS, NMR is not discriminatory toward certain metabolites. Its sensitivity is not dependent upon the chemical polarity of metabolites and, as such, can be used for broad metabolome coverage.[97] McClay et al. used ^{1}H NMR as a metabolomics screening tool for the identification of plasma and urinary biomarkers for chronic obstructive pulmonary

disease in addition to identifying metabolites associated with baseline lung function.[98] Samuelsson et al. used NMR-based metabolomics to study metabolic responses of rainbow trout exposed to the synthetic contraceptive estrogen ethinylestradiol.[99] Despite the advantages of NMR, it is still a less sensitive detection method than MS. NMR is able to detect nanogram quantities, while MS has detection limits ranging from femtograms to pictograms.[100-102] Furthermore, the complexity of biological samples leads to major spectral overlap of NMR signals from many metabolites, especially in the aliphatic region. ^{1}H *J*-resolved NMR spectroscopy and two-dimensional methods (e.g. correlation spectroscopy, total correlation spectroscopy, or heteronuclear single quantum conherence spectroscopy) have been developed to increase the resolution and sensitivity of NMR analysis. *J*-resolved NMR separates metabolite chemical shifts and *J*-coupling into two dimensions, thereby, reducing spectral congestion and increasing metabolite specificity.[103] Viant demonstrated how *J*-resolved NMR increases the amount of extractable metabolic information from NMR spectra through the investigation of embryogenesis in an established fish model for developmental toxicology.[104] Metabolomics studies also combine both NMR and MS approaches to broaden metabolome coverage as these detection platforms can complement each other. ^{1}H NMR and UPLC-MS analyses have been used to determine the geographical origins of differing herbal medicines and to identify the primary and secondary metabolites responsible for the discrimination.[105]

1.4 Biological Applications of Metabolic Fingerprinting

Metabolomics has been used for a wide variety of biological applications. Those highlighted in Sections 1.4.1 and 1.4.2 are directly related to the work presented in this dissertation.

1.4.1 Metabolomics in Oncology (Oncometabolomics)

Cancer is currently the 2nd leading cause of death in the United States but is expected to surpass heart disease as the leading cause within the next few years.[106, 107] In the US alone, 1 in 4 deaths is due to cancer.[106] Although the cancer death rate has been declining for nearly 2 decades, more than 1.6 million new cancer cases and 589,430 cancer deaths are estimated to occur in 2015.[107] Although early diagnosis leads to improved prognosis, there is a knowledge deficit regarding the biochemical processes of the origin and proliferative molecular mechanisms of many types of cancers which has also led to a shortage of sensitive and specific cancer biomarkers.[108, 109] Researchers have looked to proteins for biomarker discovery for over a century. Yet, less than 10 proteins have progressed to FDA-approved cancer diagnostic tests.[109] Recently, metabolomics has become more attractive for cancer research due to the development of technologies that can discriminate metabolic fingerprints among healthy, precancerous, and cancerous cells or tissues.[110] Modern interest in oncometabolomics stemmed from the late 1980s claim that cancer could be identified by NMR spectra of blood samples;[111] however, this study was found to be falsified and consequently tainted the field of metabolomics.[112] Yet, the notion that metabolomics could aid in understanding and detecting cancer persisted.[112] Changes in metabolite levels occur after modifications of the genome and proteome linked to cancer-associated biochemical reactions, yet still before cell morphologic

changes associated with cancer.[113] Therefore, examining the metabolome is appealing for early diagnostics of cancer since this analysis should detect the initial stages of carcinogenesis, as well as the observation of cancerous effects and subsequent therapeutic intervention.[55] Moreover, metabolomics could aid in the understanding of the mechanisms of cancer development and proliferation.

While traditional biomarker discovery studies focused on finding one marker for disease, biomarker panels which are typically sought after in metabolomics research have the potential to give better sensitivity and/or specificity than single biomarkers alone. It has been shown, for example, that combinations of known bladder cancer biomarkers can increase diagnostic sensitivity on average to 91 – 98%.[114] For brain cancer, Florian et al. demonstrated that *in vitro* NMR and HPLC metabolic profiles could distinguish between three types of human brain and nervous system tumors (meningiomas, neuroblastomas, and glioblastomas) using biomarker panels.[115] Discriminant metabolites from NMR analysis included alanine, glutamate, creatine, phosphorylcholine, and threonine, while HPLC analysis showed that taurine, γ-aminobutyric acid, and serine discriminated among the 3 tumor types. These early findings demonstrated that metabolomics has potential for the development of brain cancer tumor lineage diagnosis.[115] Maxwell et al. found that ^{1}H NMR metabolomic profiles had 85% accuracy when used to differentiate meningiomas from other brain tumors.[116] This accuracy was increased to 89% when only creatine and glutamine signals from the spectra were used for classification. Additionally, inositol levels significantly correlated with glioma grade. For colorectal cancer, Qiu et al. found that 5 metabolites associated with perturbations of glycolysis, arginine and proline metabolism, fatty acid metabolism, and oleamide metabolism were linked to colorectal

cancer morbidity using GC-MS and UPLC-MS serum metabolomic profiles.[117] Nishiumi et al. employed GC-MS based serum metabolomic fingerprinting to create a prediction model for colorectal cancer based on levels of 2-hydroxybutyrate, aspartic acid, kynurenine, and cystamine.[118] Validation of the prediction model with test samples resulted in 83.1%, 81.0%, and 82.0% sensitivity, specificity, and accuracy, respectively. The model also detected early-stage colorectal cancer with 82.8% sensitivity.

Metabolomics is actively used to understand the molecular mechanisms by which chemotherapeutic drugs attack and destroy cancerous cells in addition to assessing the efficacy of treatment. Corominas-Faja et al. found that metformin-treated breast cancer cells had significant accumulation of 5-formimino-tetrahydrofolate using UPLC-MS metabolic profiles.[119] 5-formimino-tetrahydrofolate carries activated one-carbon units that are essential for the *de novo* synthesis of purines and pyrimidines. Concurrently, *de novo* synthesis of glutathione, a folate-dependent pathway involved in one-carbon metabolism, was reduced in response to metformin. Activation of the DNA repair protein ATM kinase and the metabolic tumor suppressor AMPK were not observed after metformin treatment. The findings suggested for the first time that metformin can function as an antifolate chemotherapeutic agent that induces the ATM/AMPK tumor suppressor after modifying the flow of carbon through folate-related one-carbon metabolic pathways. Choline phospholipid metabolism is elevated in many cancers, and the increase of total choline-containing metabolites may be used as a predictive biomarker for monitoring treatment efficacy in targeted therapies.[112, 120] Employing combined magnetic resonance imaging and three-dimensional spectroscopic imaging, Mueller-Lisse et al. studied the metabolic effects of hormone-deprivation therapy in

prostate cancer patients. There was a substantial time-dependent decrease of choline, creatine, citrate, and polyamines, which ultimately resulted in the complete loss of these metabolites in 25% of patients on long-term hormone-deprivation therapy. Residual prostate cancer could be detected by elevated choline levels (choline/creatine ratio ≥ 1.5). These findings provided both a measure to detect residual prostate cancer and a time-course of metabolic response following hormone-deprivation therapy.[121] Metabolomics is also used to determine molecular targets for new chemotherapies, specifically focusing on perturbed pathways involved in cancer growth, proliferation, and metastasis.[112] Folger et al. developed a flux balance analysis model of cancer metabolism that captured the key metabolic alterations that occur across many cancer types—the first computational approach for interpreting the ever-growing body of metabolomics and proteomics data.[122] The model was used to predict 52 cytostatic chemotherapeutic targets, of which 60% are not currently targeted by known, approved, or experimental anticancer drugs. Potential selective treatments for specific cancers that depend on cancer type-specific down-regulation of gene expression and somatic mutations were also presented.

1.4.2 Metabolomics in Chemical Ecology (Ecometabolomics)

Chemical ecology is concerned with the chemical signals that organisms produce and of which they respond that lead to interactions amongst themselves. These chemically-mediated interactions have allowed researchers to become more knowledgeable of how organisms locate food and habitats, as well as avoid predators and pathogens, compete with each other, and mate.[123, 124] These interactions can dramatically influence ecosystem structure.[124, 125] Chemical ecology advancements have correlated to advancements in the sensitivities of analytical technologies which have allowed for the

discovery of many compounds utilized for chemical signaling.[123, 124] The traditional analysis technique for the discovery of compounds used for chemical signaling is bioassay-guided fractionation. While bioassay-guided fractionation can be powerful, bioactivity resulting from unstable compounds cannot be traced back to its original molecular source. Furthermore, compounds that interact synergistically are not detected since individual fractions may have no effect on bioassays even though combined fractions may illicit a biochemical response.[124] Bioassay-guided fractionation is also time consuming and not high-throughput. Although studies using metabolomics to gain knowledge regarding chemically-mediated interactions are scarce,[1] researchers have begun to exploit the advantages that metabolomics has over the traditional bioassay-guided fractionation. As all metabolites are measured at once, unstable compounds are more likely to be detected.[1, 124] Moreover, metabolomic investigations can evaluate the different responses a particular species phenotype may have in response to chemically-mediated interactions or environmental changes in addition to the metabolic pathways involved in those responses.[1] Metabolomics has been used to explore the effect of drought on the growth on *Oryza sativa*,[126] the global responses of *Arabidopsis thaliana* to nutritional stresses,[127] and the impact of pollution exposure on the metabolism of *Mytilus galloprovincialis*.[128] Metabolomics has been used to investigate the impact of biotic factors on organisms as well. Choi et al. found 5-caffeoylquinic acid, α-linolenic acid analogues, and sesqui- and diterpenoids were connected to the systemic acquired resistance of tobacco to tobacco mosaic virus. Before their metabolomics study, only genes and proteins involved in the systemic acquired resistance had been characterized.[129] Schroeder et al. found that pinoresinol, a plant lignan, is a minor component in the

defensive secretion produced by the glandular hairs of *Pieris rapae*, cabbage butterfly caterpillars.[130] The caterpillars obtain pinoresinol, which is a deterrant to ants, from their food source—cabbage (*Brassica oleracea*). Pinoresinol proved to be more potent than mayolene-16, a primary component belonging to the mayolene group of lipids previously reported to be responsible for the action of the defensive secretion. Pieiris et al. used metabolomics to study the chemically-mediated interactions of the wood decay basidiomycete fungus *Stereum hirsutum* with its competitors *Coprinus micaceus* and *Coprinus disseminates*.[131] There were increased levels of 3 metabolites including malic acid and 1,2-dihydroxyanthraquinone when *S. hirsutum* was overgrown by *C. disseminates*. When *S. hirsutum* interacted with *C. micaceus*, decreased levels of 7 metabolites including 2-methyl-2, 3-dihydroxypropionic acid and pyridoxine were observed.

1.5 Traditional Mass Spectrometry-Based Untargeted Metabolomics Workflow

Traditional MS-based metabolomics research is performed using chromatography coupled to MS and results in complex three-dimensional data sets. A traditional MS-based untargeted metabolomics workflow is shown in Figure 1.3. The steps involved are further described in Sections 1.5.1 – 1.5.5.

1.5.1 Experimental Design

Untargeted metabolomics is an impartial, all-inclusive approach for detecting as many metabolites as possible without *a priori* knowledge of their identity. These analyses are the first steps toward designing targeted assays to study specific metabolic pathways,

detect clinical disease, or monitor environmental phenomena; thus, the experimental design is of the utmost importance.[132] Experimental details that should be established early include predetermined sample classes (e.g., cancer vs. non-cancer or exposed vs. control) in addition to the types of biological samples needed and the number of samples to be included in the study. Instrumentation needed for analyses should also be determined since its analytical sensitivity can influence sample preparation methods. An assessment of analytical sensitivity and repeatability during metabolomics studies is obtained through the utilization of quality control (QC) samples.[133]

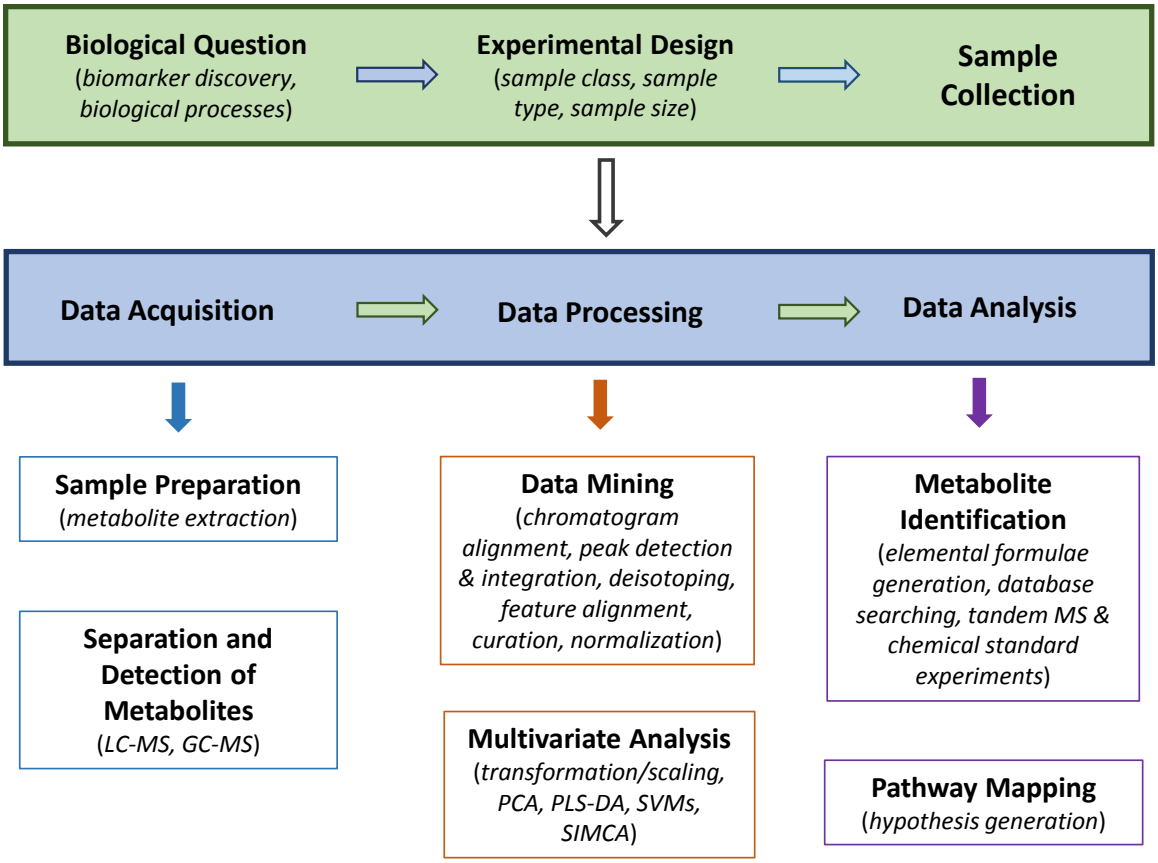

Figure 1.3: A typical untargeted metabolomics workflow.

While QCs can vary, it has been recommended that a pooling of all samples to be analyzed serve as the QC for metabolomics studies.[134] Furthermore, experimental details must consider possible confounding factors such as age, gender, and ethnicity, to ensure that resultant multivariate analysis is not influenced by such variables.[135, 136] Analysis methods should have limited technical variation; if technical variation exceeds the biological variation of the system under investigation, results from the ensuing study may be invalid as multivariate analysis methods could solely interpret technical variation and not useful biological information. Steps taken to ensure this include conducting replicate analyses, cleaning instrumentation between sample batches, and optimizing instrumental parameters before beginning any metabolomics study.[134]

1.5.2 Biological Sample Preparation

Many types of biological samples are used for metabolomics studies—serum,[40, 77] plasma,[40, 99] urine,[40, 64, 87] cerebrospinal fluid,[12] exhaled breath condensate,[137] and tissue.[63, 75] As these sample matrices also contain proteins, the primary sample preparation method involved in metabolomics research is extraction of metabolites. Metabolite extraction is challenging due to the chemical diversity of the metabolome.[100] Typically, organic solvents are used to precipitate proteins by disrupting hydrogen bonding to water and, subsequently, extract soluble metabolites. However, these extraction methods may deplete certain metabolite levels, i.e. protein-bound metabolites, and suffer from matrix interference.[100, 138] The solvent system used for metabolite extraction greatly impacts the type of metabolites detected during metabolomic analyses—polar solvents will extract polar metabolites and a nonpolar solvent system will be biased towards extracting nonpolar metabolites. Moreover, the solvent system

used should be able to reproducibly extract metabolites to ensure minimum technical variation in the resultant data. Testing 14 different metabolite extraction methods for reversed-phase LC-MS serum metabolomics, Want et al. found that utilizing methanol as the extraction solvent was the most effective, reproducible metabolite extraction method.[139] The resultant serum extracts contained over 2,000 detected metabolic features and less than 2% residual protein. Furthermore, methanol extracted metabolites with varying degrees of hydrophobicity. Interestingly, combining all of the metabolite extraction methods resulted in the detection of over 10,000 unique metabolic features, thereby demonstrating that metabolite extraction procedures can bias the types of detected metabolites. Using 12 protocols to extract intracellular metabolites from Chinese hamster ovary cells, Dietmair et al. found that the concentration of extracted intracellular metabolites was highest with a cold 50:50 (v/v) mixture of acetonitrile and water.[140] Additionally, this method was able to fully extract/recover all analytes of an experimental chemical standard metabolite mixture. Metabolites can also be fractionated after proteins are eliminated; liquid-liquid extraction can be performed to further separate extracted metabolites into polar aqueous and lipophilic organic fractions.[4] Masson et al. demonstrated that UPLC-MS metabolic profiling of polar and nonpolar liver tissue metabolite extracts benefitted from aqueous extraction with methanol/water followed by an organic extraction with dichloromethane/methanol; although, dried extracts had to be resuspended in methanol/water before UPLC-MS analysis.[141] The median coefficients of variation for metabolic feature intensities of aqueous extracts was <20%, while it was <30% for organic extracts. As the primary aim of untargeted metabolomics is to detect as many metabolites as possible without *a priori* knowledge of their identity, metabolite

extraction protocols should be optimized before beginning a study to ensure broad metabolome coverage.

1.5.3 Data Analysis and Metabolic Feature Extraction

Resultant untargeted metabolomics data files contain three dimensions: (1) chromatographic retention time, (2) mass-to-charge ratio (m/z), and (3) abundance. The acquired data sets are mined to extract metabolic features (unique retention time, m/z pairs) across all analyzed samples. Metabolic feature lists can contain hundreds to thousands of features.[142] The most commonly used open-source software tools to accomplish this are MZmine[143] and XCMS,[144] but vendor supplied software also exists for this purpose. These software tools support many MS data file types. Furthermore, they can be utilized for data visualization. Data mining procedures generally include chromatogram alignment to eliminate retention time variation across analyses that may result from sample carryover or column degradation when conducting large studies. After chromatograms are aligned, peaks are detected and integrated so that peak areas can be extracted for all metabolic features that pass chromatographic and spectral noise thresholds and chromatographic peak width constraints.[145] The resultant metabolic feature list is deisotoped and adduct ions may be removed to ensure that metabolic features are not represented more than once. Feature lists are aligned to match metabolic features across all data files. MZmine allows users to "gap-fill" the peak list to ensure that none of the algorithms used for the above process disregard metabolic features that are actually present in the raw data.[146] Typically, metabolic feature lists are curated to eliminate any features that appear in blank samples and those that are not present in a user-specified percentage of the real samples so that the metabolic features are

representative of the samples included in the study. Lastly, the peak list is normalized. Sample data is commonly normalized to the sum of its feature abundances.[145]

1.5.4 Pattern Recognition by Multivariate Analysis

From a statistical point of view, analysis of metabolomic data sets represents a significant challenge, and robust approaches are necessary to handle, extract and classify the relevant information from the vast amount of data generated.[147] Data sets can be explored through principal components analysis (PCA). PCA interprets and transforms the metabolic features into a small number of principal components that capture maximal variance in the data.[142] Visualization of this variance assists in the discernment of data trends.[4] As such, PCA allows the user to determine if technical variance, such as instrumental drift which leads to batch effects or sample carryover, will affect the analysis of biological variance, in addition to pointing out sample outliers. Metabolite concentration differences in metabolomics data sets can be as high as 5000-fold.[142] Multivariate analysis techniques are sensitive to large metabolite abundances and concentration differences even if they are not biologically relevant. As such, metabolomics data is transformed or scaled to reduce large abundances so that low abundances are not dwarfed and multivariate statistical analysis is more prone to identify pertinent biological changes.[148] Different scaling or transformation methods used include mean-centering, autoscaling, Pareto scaling, and general logarithm transformation.[149] Supervised classification techniques such as partial least squares discriminant analysis (PLS-DA), soft-independent method of class analogy (SIMCA),[150] and support vector machines (SVMs)[151] are generally used to discriminate sample classes and assist in the interpretation of the biological information gained from metabolomics studies.[152] These

supervised techniques incorporate sample class membership into their analysis, so that biological patterns in feature abundances that discriminate sample classes can be known. As this modeling is done with metabolic feature lists that may contain thousands of features, it is important to couple the statistical modeling with variable selection methods that extract the most important metabolic features needed for classification. Examples of variable selection methods are variable importance in projection (VIP) scores,[151] recursive feature elimination (RFE),[153] genetic algorithms,[132] sensitivity ratios,[154] and S-plots.[155] Variable selection also allows metabolite identification to become more manageable and eliminates noisy metabolic features that do not contribute to the biological variation that aids in classification. Oftentimes in untargeted metabolomics studies, a more global assessment of perturbed metabolism is sought after, and metabolic features are analyzed for statistically significant differences between sample groupings. This can be accomplished by nonparametric statistical methods, analysis of variance (ANOVA), and a Student's *t* Test.[142]

1.5.5 Metabolite Identification and Pathway Mapping

Metabolites responsible for class differentiation/discrimination are identified so that hypotheses regarding pathway perturbations can be made or identities of potential biomarkers can be known. Many MS-based metabolomics studies employ high-resolution instrumentation capable of providing accurate masses for metabolites. Generated elemental formulae based on the exact mass and isotopic patterns of the metabolites can be searched against metabolite databases,[156] such as HMDB,[46] METLIN,[48] and the LIPID MAPS database.[50] Despite the growth of metabolite databases over the years, a substantial number of metabolic features are still not matched to metabolites.[157]

Metabolites are more confidently annotated when their fragmentation patterns and chromatographic retention times are matched to chemical standards.[156] However, depending on the MS instrumentation, tandem MS fragmentation spectra may be difficult to obtain for low concentration metabolites whose precursor ion abundance is not high enough for sensitive quadrupole selection and subsequent collision induced dissociation due to ion transmission losses.[158] If a chemical standard is not available, metabolite fragmentation patterns can be searched against the MS/MS METLIN database[48] and MassBank;[159] literature searches can also be conducted. Although tedious, sometimes fragmentation patterns must be manually analyzed for metabolite identity and to discriminate between different isobaric species. Once metabolite identities are validated, databases such as KEGG[49] and MetaCyc[51] are utilized to determine the metabolic pathways in which the metabolites play a role and the upstream biological molecules to which they are linked. Hypotheses regarding the metabolic state of the sample classes can then be developed.

1.6 Conclusion

Metabolomics has a rich history in clinical chemistry but is continuing to add girth to many additional biological fields. As metabolite distributions are representative of phenotype, MS-based untargeted metabolomics analysis is capable of providing new insight into disease detection and progression as well as increasing our knowledge of ecological interactions in particular environmental systems. Still, shortcomings in metabolomics have been realized and advances in systems biology data integration and high-throughput analytical technologies are needed. Pairing metabolomic information with that acquired from other "omics" disciplines gives a more all-inclusive snapshot of

perturbations to metabolic pathways since it is difficult to determine causal relationships within metabolic networks with only metabolomic data. Additionally, chromatographic separation methods coupled to MS are very powerful for metabolome analysis, but they are not without challenges for large metabolomics studies—lengthy analysis times—so there is still a need for developing higher throughput MS approaches.

1.7 References

(1) Sardans, J.; Penuelas, J.; Rivas-Ubach, A., Ecological metabolomics: overview of current developments and future challenges. *Chemoecology* **2011,** *21* (4), 191-225.

(2) Dunn, W. B.; Broadhurst, D. I.; Atherton, H. J.; Goodacre, R.; Griffin, J. L., Systems level studies of mammalian metabolomes: the roles of mass spectrometry and nuclear magnetic resonance spectroscopy. *Chem. Soc. Rev.* **2011,** *40* (1), 387-426.

(3) Griffiths, W. J.; Koal, T.; Wang, Y.; Kohl, M.; Enot, D. P.; Deigner, H.-P., Targeted metabolomics for biomarker discovery. *Angew. Chem. Int. Ed.* **2010,** *49* (32), 5426-5445.

(4) Dettmer, K.; Aronov, P. A.; Hammock, B. D., Mass spectrometry-based metabolomics. *Mass Spectrom. Rev.* **2007,** *26* (1), 51-78.

(5) van der Greef, J.; Smilde, A. K., Symbiosis of chemometrics and metabolomics: past, present, and future. *J. Chemom.* **2005,** *19* (5-7), 376-386.

(6) Nicholson, J. K.; Lindon, J. C., Systems biology: metabonomics. *Nature* **2008,** *455* (7216), 1054-1056.

(7) Dalgliesh, C. E., Two-dimensional paper chromatography of urinary indoles and related substances. *Biochem. J.* **1956,** *64* (3), 481.

(8) Williams, R. J., *Biochemical Institute Studies IV. Individual Metabolic Patterns and Human Disease: An Exploratory Study Utilizing Predominantly Paper Chromatographic Methods, U. Texas Publication No. 5109*. University of Texas: Austin, TX, 1951.

(9) Gates, S. C.; Sweeley, C. C., Quantitative metabolic profiling based on gas chromatography. *Clin. Chem.* **1978,** *24* (10), 1663-1673.

(10) Young, D. S., High pressure column chromatography of carbohydrates in the clinical laboratory. *Am. J. Clin. Pathol.* **1970,** *53* (5), 803-810.

(11) Williams, G. Z.; Young, D. S.; Stein, M. R.; Cotlove, E., Biological and analytic components of variation in long-term studies of serum constituents in normal subjects I. Objectives, subject selection, laboratory procedures, and estimation of analytic deviation. *Clin. Chem.* **1970,** *16* (12), 1016-1021.

(12) Malcolm, R. D.; Leonards, R., Gas-liquid chromatographic profile of neutral and acidic metabolites in cerebrospinal fluid from newborns and infants. *Clin. Chem.* **1976,** *22* (5), 623-626.

(13) Scott, C. D.; Chilcote, D. D.; Katz, S.; Pitt, W. W., Advances in the application of high resolution liquid chromatography to the separation of complex biological mixtures. *J. Chromatogr. Sci.* **1973,** *11* (2), 96-100.

(14) Pauling, L.; Robinson, A. B.; Teranishi, R.; Cary, P., Quantitative analysis of urine vapor and breath by gas-liquid partition chromatography. *Proc. Natl. Acad. Sci. USA* **1971,** *68* (10), 2374-2376.

(15) Dempster, A. J., A new method of positive ray analysis. *Phys. Rev.* **1918,** *11* (4), 316-325.

(16) Aston, F. W., LXXIV. A positive ray spectrograph. *Philos. Mag.* **1919,** *38* (228), 707-714.

(17) Horning, E. C.; Horning, M. G., Metabolic profiles: gas-phase methods for analysis of metabolites. *Clin. Chem.* **1971,** *17* (8), 802-809.

(18) Horning, E. C.; Horning, M. G., Human metabolic profiles obtained by GC and GC/MS. *J. Chromatogr. Sci.* **1971,** *9* (3), 129-140.

(19) Yamashita, M.; Fenn, J. B., Electrospray ion source. Another variation on the free-jet theme. *J. Phys. Chem.* **1984,** *88* (20), 4451-4459.

(20) Carroll, D.; Dzidic, I.; Stillwell, R.; Haegele, K.; Horning, E., Atmospheric pressure ionization mass spectrometry. Corona discharge ion source for use in a liquid chromatograph-mass spectrometer-computer analytical system. *Anal. Chem.* **1975,** *47* (14), 2369-2373.

(21) Erni, F., Liquid chromatography mass spectrometry in the pharmaceutical industry: objectives and needs. *J. Chromatogr. A* **1982,** *251* (2), 141-151.

(22) Henion, J. D., Continuous monitoring of total micro LC eluant by direct liquid introduction LC/MS. *J. Chromatogr. Sci.* **1981,** *19* (2), 57-64.

(23) Henion, J. D., Drug analysis by continuously monitored liquid chromatography/mass spectrometry with a quadrupole mass spectrometer. *Anal. Chem.* **1978,** *50* (12), 1687-1693.

(24) Covey, T. R.; Crowther, J. B.; Dewey, E. A.; Henion, J. D., Thermospray liquid chromatography/mass spectrometry determination of drugs and their metabolites in biological fluids. *Anal. Chem.* **1985,** *57* (2), 474-481.

(25) Games, D. E.; Alcock, N. J.; van der Greef, J.; Nyssen, L. M.; Maarse, H.; de Brauw, M. C. T. N., Analysis of pepper and capsicum oleoresins by high-performance

liquid chromatography—mass spectrometry and field desorption mass spectrometry. *J. Chromatogr. A* **1984,** *294*, 269-279.

(26) Politzer, I. R.; Dowty, B. J.; Laseter, J. L., Use of gas chromatography and mass spectrometry to analyze underivatized volatile human or animal constituents of clinical interest. *Clin. Chem.* **1976,** *22* (11), 1775-1788.

(27) Mayron, L. W.; Kaplan, E.; Alling, S.; Becktel, J., Drug-abuse and control populations differentiated by a laboratory profile. *Clin. Chem.* **1974,** *20* (2), 172-176.

(28) Rhodes, G.; Miller, M.; McConnell, M. L.; Novotny, M., Metabolic abnormalities associated with diabetes mellitus, as investigated by gas chromatography and pattern-recognition analysis of profiles of volatile metabolites. *Clin. Chem.* **1981,** *27* (4), 580-585.

(29) McConnell, M. L.; Rhodes, G.; Watson, U.; Novotný, M., Application of pattern recognition and feature extraction techniques to volatile constituent metabolic profiles obtained by capillary gas chromatography. *J. Chromatogr. B Biomed. Appl.* **1979,** *162* (4), 495-506.

(30) Blomquist, G.; Johansson, E.; Söderström, B.; Wold, S., Classification of fungi by means of pyrolysis-gas chromatography-pattern recognition. *J. Chromatogr. A* **1979,** *173* (1), 19-32.

(31) van der Greef, J.; Tas, A. C.; Bouwman, J.; de Brauw, M. C. T. N.; Schreurs, W. H. P., Evaluation of field-desorption and fast atom-bombardment mass spectrometric profiles by pattern recognition techniques. *Anal. Chim. Acta* **1983,** *150*, 45-52.

(32) Tas, A. C.; Van der Greef, J., Mass spectrometric profiling and pattern recognition. *Mass Spectrom. Rev.* **1994,** *13* (2), 155-181.

(33) Purcell, E. M.; Torrey, H.; Pound, R. V., Resonance absorption by nuclear magnetic moments in a solid. *Phys. Rev.* **1946,** *69* (1-2), 37.

(34) Bloch, F.; Hansen, W. W.; Packard, M., Nuclear induction. *Phys. Rev.* **1946,** *69* (3-4), 127-127.

(35) Nicholson, J. K.; Wilson, I. D., High resolution proton magnetic resonance spectroscopy of biological fluids. *Prog. Nucl. Magn. Reson. Spectrosc.* **1989,** *21* (4), 449-501.

(36) Bell, J. D.; Brown, J. C. C.; Sadler, P. J., NMR studies of body fluids. *NMR Biomed.* **1989,** *2* (5-6), 246-256.

(37) Iles, R. A.; Hind, A. J.; Chalmers, R. A., Use of proton nuclear magnetic resonance spectroscopy in detection and study of organic acidurias. *Clin. Chem.* **1985,** *31* (11), 1795-1801.

(38) Williams, S. R.; Iles, R. A.; Chalmers, R. A., Spin-echo and 2-dimensional 1 H nuclear magnetic resonance studies on urinary metabolites from patients with 2-methylacetocetyl CoA thiolase deficiency. *Clin. Chim. Acta* **1986,** *159* (2), 153-161.

(39) Yamaguchi, S.; Koda, N.; Eto, Y.; Aoki, K., Rapid screening of metabolic disease by proton NMR urinalysis. *Lancet* **1984,** *324* (8397), 284.

(40) Nicholson, J. K.; O'Flynn, M. P.; Sadler, P. J.; Macleod, A. F.; Juul, S. M.; Sonksen, P. H., Proton-nuclear-magnetic-resonance studies of serum, plasma and urine from fasting normal and diabetic subjects. *Biochem. J.* **1984,** *217*, 365-375.

(41) Soga, T.; Heiger, D. N., Amino acid analysis by capillary electrophoresis electrospray ionization mass spectrometry. *Anal. Chem.* **2000,** *72* (6), 1236-1241.

(42) Soga, T.; Ohashi, Y.; Ueno, Y.; Naraoka, H.; Tomita, M.; Nishioka, T., Quantitative metabolome analysis using capillary electrophoresis mass spectrometry. *J. Proteome Res.* **2003,** *2* (5), 488-494.

(43) Plumb, R.; Castro-Perez, J.; Granger, J.; Beattie, I.; Joncour, K.; Wright, A., Ultra-performance liquid chromatography coupled to quadrupole-orthogonal time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **2004,** *18* (19), 2331-2337.

(44) Plumb, R. S.; Granger, J. H.; Stumpf, C. L.; Johnson, K. A.; Smith, B. W.; Gaulitz, S.; Wilson, I. D.; Castro-Perez, J., A rapid screening approach to metabonomics using UPLC and oa-TOF mass spectrometry: application to age, gender and diurnal variation in normal/Zucker obese rats and black, white and nude mice. *Analyst* **2005,** *130* (6), 844-849.

(45) Wishart, D. S.; Tzur, D.; Knox, C.; Eisner, R.; Guo, A. C.; Young, N.; Cheng, D.; Jewell, K.; Arndt, D.; Sawhney, S., HMDB: the human metabolome database. *Nucleic Acids Res.* **2007,** *35* (suppl 1), D521-D526.

(46) Wishart, D. S.; Jewison, T.; Guo, A. C.; Wilson, M.; Knox, C.; Liu, Y.; Djoumbou, Y.; Mandal, R.; Aziat, F.; Dong, E., HMDB 3.0—the human metabolome database in 2013. *Nucleic Acids Res.* **2012,** *41* (suppl 1), D801-D807.

(47) Wishart, D. S.; Knox, C.; Guo, A. C.; Eisner, R.; Young, N.; Gautam, B.; Hau, D. D.; Psychogios, N.; Dong, E.; Bouatra, S., HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Res.* **2009,** *37* (suppl 1), D603-D610.

(48) Smith, C. A.; O'Maille, G.; Want, E. J.; Qin, C.; Trauger, S. A.; Brandon, T. R.; Custodio, D. E.; Abagyan, R.; Siuzdak, G., METLIN: a metabolite mass spectral database. *Ther. Drug Monit.* **2005,** *27* (6), 747-751.

(49) Kanehisa, M.; Goto, S., KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **2000,** *28* (1), 27-30.

(50) Sud, M.; Fahy, E.; Cotter, D.; Brown, A.; Dennis, E. A.; Glass, C. K.; Merrill, A. H.; Murphy, R. C.; Raetz, C. R.; Russell, D. W., Lmsd: lipid maps structure database. *Nucleic Acids Res.* **2007,** *35* (suppl 1), D527-D532.

(51) Caspi, R.; Foerster, H.; Fulcher, C. A.; Kaipa, P.; Krummenacker, M.; Latendresse, M.; Paley, S.; Rhee, S. Y.; Shearer, A. G.; Tissier, C., The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res.* **2008,** *36* (suppl 1), D623-D631.

(52) Dunn, W. B.; Ellis, D. I., Metabolomics: current analytical platforms and methodologies. *TrAC, Trends Anal. Chem.* **2005,** *24* (4), 285-294.

(53) Ruzsanyi, V.; Baumbach, J. I.; Sielemann, S.; Litterst, P.; Westhoff, M.; Freitag, L., Detection of human metabolites using multi-capillary columns coupled to ion mobility spectrometers. *J. Chromatogr. A* **2005,** *1084* (1), 145-151.

(54) Shurubor, Y. I.; Matson, W. R.; Willett, W. C.; Hankinson, S. E.; Kristal, B. S., Biological variability dominates and influences analytical variance in HPLC-ECD studies of the human plasma metabolome. *BMC Clin. Pathol.* **2007,** *7* (1), 9.

(55) Putri, S. P.; Yamamoto, S.; Tsugawa, H.; Fukusaki, E., Current metabolomics: technological advances. *J. Biosci. Bioeng.* **2013,** *116* (1), 9-16.

(56) Gomase, V.; Changbhale, S.; Patil, S.; Kale, K., Metabolomics. *Curr. Drug Metab.* **2008,** *9* (1), 89-98.

(57) Halket, J. M.; Waterman, D.; Przyborowska, A. M.; Patel, R. K.; Fraser, P. D.; Bramley, P. M., Chemical derivatization and mass spectral libraries in metabolic profiling by GC/MS and LC/MS/MS. *J. Exp. Bot.* **2005,** *56* (410), 219-243.

(58) Jonsson, P.; Gullberg, J.; Nordström, A.; Kusano, M.; Kowalczyk, M.; Sjöström, M.; Moritz, T., A strategy for identifying differences in large series of metabolomic samples analyzed by GC/MS. *Anal. Chem.* **2004,** *76* (6), 1738-1745.

(59) Roessner, U.; Wagner, C.; Kopka, J.; Trethewey, R. N.; Willmitzer, L., Simultaneous analysis of metabolites in potato tuber by gas chromatography–mass spectrometry. *Plant J.* **2000,** *23* (1), 131-142.

(60) Little, J. L., Artifacts in trimethylsilyl derivatization reactions and ways to avoid them. *J. Chromatogr. A* **1999,** *844* (1), 1-22.

(61) Lee, A. L.; Bartle, K. D.; Lewis, A. C., A model of peak amplitude enhancement in orthogonal two-dimensional gas chromatography. *Anal. Chem.* **2001,** *73* (6), 1330-1335.

(62) Habram, M.; Welsch, T., Improvement of the detection limit in capillary gas Chromatographic trace analysis of C2–C4 hydrocarbons by post column cryogenic trapping. *J. High. Resolut. Chromatogr.* **1999,** *22* (6), 335-338.

(63) Welthagen, W.; Shellie, R. A.; Spranger, J.; Ristow, M.; Zimmermann, R.; Fiehn, O., Comprehensive two-dimensional gas chromatography–time-of-flight mass spectrometry (GC× GC-TOF) for high resolution metabolomics: biomarker discovery on spleen tissue extracts of obese NZO compared to lean C57BL/6 mice. *Metabolomics* **2005,** *1* (1), 65-73.

(64) Kouremenos, K. A.; Pitt, J.; Marriott, P. J., Metabolic profiling of infant urine using comprehensive two-dimensional gas chromatography: application to the diagnosis of organic acidurias and biomarker discovery. *J. Chromatogr. A* **2010,** *1217*, 104-111.

(65) Mohler, R. E.; Dombek, K. M.; Hoggard, J. C.; Young, E. T.; Synovec, R. E., Comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry analysis of metabolites in fermenting and respiring yeast cells. *Anal. Chem.* **2006,** *78* (8), 2700-2709.

(66) Alpert, A. J., Hydrophilic-interaction chromatography for the separation of peptides, nucleic acids and other polar compounds. *J. Chromatogr. A* **1990,** *499*, 177-196.

(67) Tolstikov, V. V.; Fiehn, O., Analysis of highly polar compounds of plant origin: combination of hydrophilic interaction chromatography and electrospray ion trap mass spectrometry. *Anal. Biochem.* **2002,** *301* (2), 298-307.

(68) Luo, B.; Groenke, K.; Takors, R.; Wandrey, C.; Oldiges, M., Simultaneous determination of multiple intracellular metabolites in glycolysis, pentose phosphate pathway and tricarboxylic acid cycle by liquid chromatography–mass spectrometry. *J. Chromatogr. A* **2007,** *1147* (2), 153-164.

(69) Whitehouse, C. M.; Dreyer, R. N.; Yamashita, M.; Fenn, J. B., Electrospray interface for liquid chromatographs and mass spectrometers. *Anal. Chem.* **1985,** *57* (3), 675-679.

(70) Ikonomou, M. G.; Blades, A. T.; Kebarle, P., Investigations of the electrospray interface for liquid chromatography/mass spectrometry. *Anal. Chem.* **1990,** *62* (9), 957-967.

(71) Wang, X.; Sun, H.; Zhang, A.; Wang, P.; Han, Y., Ultra-performance liquid chromatography coupled to mass spectrometry as a sensitive and powerful technology for metabolomic studies. *J. Sep. Sci.* **2011,** *34* (24), 3451-3459.

(72) Zhang, A.; Sun, H.; Wang, P.; Han, Y.; Wang, X., Modern analytical techniques in metabolomics analysis. *Analyst* **2012,** *137* (2), 293-300.

(73) Evans, A. M.; DeHaven, C. D.; Barrett, T.; Mitchell, M.; Milgram, E., Integrated, nontargeted ultrahigh performance liquid chromatography/electrospray ionization tandem mass spectrometry platform for the identification and relative quantification of the small-molecule complement of biological systems. *Anal. Chem.* **2009,** *81* (16), 6656-6667.

(74) Wilson, I. D.; Plumb, R.; Granger, J.; Major, H.; Williams, R.; Lenz, E. M., HPLC-MS-based methods for the study of metabonomics. *J. Chromatogr. B* **2005,** *817* (1), 67-76.

(75) Want, E. J.; Masson, P.; Michopoulos, F.; Wilson, I. D.; Theodoridis, G.; Plumb, R. S.; Shockcor, J.; Loftus, N.; Holmes, E.; Nicholson, J. K., Global metabolic profiling of animal and human tissues via UPLC-MS. *Nat. Protoc.* **2013,** *8* (1), 17-32.

(76) Zhang, A.; Sun, H.; Han, Y.; Yan, G.; Wang, X., Urinary metabolic biomarker and pathway study of hepatitis B virus infected patients based on UPLC-MS system. *PloS One* **2013,** *8* (5), e64381.

(77) Gonzalez, E.; van Liempd, S.; Conde-Vancells, J.; Gutierrez-de Juan, V.; Perez-Cormenzana, M.; Mayo, R.; Berisa, A.; Alonso, C.; Marquez, C. A.; Barr, J., Serum UPLC-MS/MS metabolic profiling in an experimental model for acute-liver injury reveals potential biomarkers for hepatotoxicity. *Metabolomics* **2012,** *8* (6), 997-1011.

(78) Ramautar, R.; Mayboroda, O. A.; Somsen, G. W.; de Jong, G. J., CE-MS for metabolomics: developments and applications in the period 2008–2010. *Electrophoresis* **2011,** *32* (1), 52-65.

(79) Smith, R. D.; Olivares, J. A.; Nguyen, N. T.; Udseth, H. R., Capillary zone electrophoresis-mass spectrometry using an electrospray ionization interface. *Anal. Chem.* **1988,** *60* (5), 436-441.

(80) Lee, E. D.; Mück, W.; Henion, J. D.; Covey, T. R., Liquid junction coupling for capillary zone electrophoresis/ion spray mass spectrometry. *Biomed. Environ. Mass Spectrom.* **1989,** *18* (9), 844-850.

(81) Issaq, H. J.; Janini, G. M.; Chan, K. C.; Veenstra, T. D., Sheathless electrospray ionization interfaces for capillary electrophoresis–mass spectrometric detection: advantages and limitations. *J. Chromatogr. A* **2004,** *1053* (1), 37-42.

(82) Garcia-Perez, I.; Vallejo, M.; Garcia, A.; Legido-Quigley, C.; Barbas, C., Metabolic fingerprinting with capillary electrophoresis. *J. Chromatogr. A* **2008,** *1204* (2), 130-139.

(83) Simó, C.; Ibáñez, C.; Gómez-Martínez, Á.; Ferragut, J. A.; Cifuentes, A., Is metabolomics reachable? Different purification strategies of human colon cancer cells provide different CE-MS metabolite profiles. *Electrophoresis* **2011,** *32* (13), 1765-1777.

(84) Hirayama, A.; Nakashima, E.; Sugimoto, M.; Akiyama, S.-i.; Sato, W.; Maruyama, S.; Matsuo, S.; Tomita, M.; Yuzawa, Y.; Soga, T., Metabolic profiling reveals new serum biomarkers for differentiating diabetic nephropathy. *Anal. Bioanal. Chem.* **2012,** *404* (10), 3101-3109.

(85) Soga, T.; Sugimoto, M.; Honma, M.; Mori, M.; Igarashi, K.; Kashikura, K.; Ikeda, S.; Hirayama, A.; Yamamoto, T.; Yoshida, H., Serum metabolomics reveals γ-glutamyl dipeptides as biomarkers for discrimination among different forms of liver disease. *J. Hepatol.* **2011,** *55* (4), 896-905.

(86) Nemes, P.; Knolhoff, A. M.; Rubakhin, S. S.; Sweedler, J. V., Metabolic differentiation of neuronal phenotypes by single-cell capillary electrophoresis–electrospray ionization-mass spectrometry. *Anal. Chem.* **2011,** *83* (17), 6810-6817.

(87) Ramautar, R.; Nevedomskaya, E.; Mayboroda, O. A.; Deelder, A. M.; Wilson, I. D.; Gika, H. G.; Theodoridis, G. A.; Somsen, G. W.; de Jong, G. J., Metabolic profiling of human urine by CE-MS using a positively charged capillary coating and comparison with UPLC-MS. *Mol. Biosyst.* **2011,** *7* (1), 194-199.

(88) Nevedomskaya, E.; Derks, R.; Deelder, A. M.; Mayboroda, O. A.; Palmblad, M., Alignment of capillary electrophoresis–mass spectrometry datasets using accurate mass information. *Anal. Bioanal. Chem.* **2009,** *395* (8), 2527-2533.

(89) Bedair, M.; Sumner, L. W., Current and emerging mass-spectrometry technologies for metabolomics. *TrAC, Trends Anal. Chem.* **2008,** *27* (3), 238-250.

(90) Kirwan, J.; Broadhurst, D.; Davidson, R.; Viant, M., Characterising and correcting batch variation in an automated direct infusion mass spectrometry (DIMS) metabolomics workflow. *Anal. Bioanal. Chem.* **2013,** *405* (15), 5147-5157.

(91) Lin, L.; Yu, Q.; Yan, X.; Hang, W.; Zheng, J.; Xing, J.; Huang, B., Direct infusion mass spectrometry or liquid chromatography mass spectrometry for human metabonomics? A serum metabonomic study of kidney cancer. *Analyst* **2010,** *135* (11), 2970-2978.

(92) Mas, S.; Villas-Bôas, S. G.; Edberg Hansen, M.; Åkesson, M.; Nielsen, J., A comparison of direct infusion MS and GC-MS for metabolic footprinting of yeast mutants. *Biotechnol. Bioeng.* **2007,** *96* (5), 1014-1022.

(93) Takats, Z.; Wiseman, J. M.; Gologan, B.; Cooks, R. G., Mass spectrometry sampling under ambient conditions with desorption electrospray ionization. *Science* **2004,** *306* (5695), 471-473.

(94) Cody, R. B.; Laramée, J. A.; Durst, H. D., Versatile new ion source for the analysis of materials in open air under ambient conditions. *Anal. Chem.* **2005,** *77* (8), 2297-2302.

(95) Cajka, T.; Riddellova, K.; Tomaniova, M.; Hajslova, J., Ambient mass spectrometry employing a DART ion source for metabolomic fingerprinting/profiling: a powerful tool for beer origin recognition. *Metabolomics* **2011,** *7* (4), 500-508.

(96) Pan, Z.; Gu, H.; Talaty, N.; Chen, H.; Shanaiah, N.; Hainline, B. E.; Cooks, R. G.; Raftery, D., Principal component analysis of urine metabolites detected by NMR and DESI–MS in patients with inborn errors of metabolism. *Anal. Bioanal. Chem.* **2007,** *387* (2), 539-549.

(97) Pan, Z.; Raftery, D., Comparing and combining NMR spectroscopy and mass spectrometry in metabolomics. *Anal. Bioanal. Chem.* **2007,** *387* (2), 525-527.

(98) McClay, J. L.; Adkins, D. E.; Isern, N. G.; O'Connell, T. M.; Wooten, J. B.; Zedler, B. K.; Dasika, M. S.; Webb, B. T.; Webb-Robertson, B.-J.; Pounds, J. G., ^{1}H nuclear magnetic resonance metabolomics analysis identifies novel urinary biomarkers for lung function. *J. Proteome Res.* **2010,** *9* (6), 3083-3090.

(99) Samuelsson, L. M.; Förlin, L.; Karlsson, G.; Adolfsson-Erici, M.; Larsson, D. J., Using NMR metabolomics to identify responses of an environmental estrogen in blood plasma of fish. *Aquat. Toxicol.* **2006,** *78* (4), 341-349.

(100) Zhou, B.; Xiao, J. F.; Tuli, L.; Ressom, H. W., LC-MS-based metabolomics. *Mol. Biosyst.* **2012,** *8* (2), 470-481.

(101) Shockcor, J. P.; Unger, S. E.; Wilson, I. D.; Foxall, P. J.; Nicholson, J. K.; Lindon, J. C., Combined HPLC, NMR spectroscopy, and ion-trap mass spectrometry with application to the detection and characterization of xenobiotic and endogenous metabolites in human urine. *Anal. Chem.* **1996,** *68* (24), 4431-4435.

(102) Albert, K., On-line use of NMR detection in separation chemistry. *J. Chromatogr. A* **1995,** *703*, 123-147.

(103) Ludwig, C.; Viant, M. R., Two-dimensional J-resolved NMR spectroscopy: review of a key methodology in the metabolomics toolbox. *Phytochem. Anal* **2010,** *21* (1), 22-32.

(104) Viant, M. R., Improved methods for the acquisition and interpretation of NMR metabolomic data. *Biochem. Biophys. Res. Commun.* **2003,** *310* (3), 943-948.

(105) Kim, E. J.; Kwon, J.; Park, S. H.; Park, C.; Seo, Y.-B.; Shin, H.-K.; Kim, H. K.; Lee, K.-S.; Choi, S.-Y.; Ryu, D. H., Metabolite profiling of *Angelica gigas* from different geographical origins using ^{1}H NMR and UPLC-MS analyses. *J. Agric. Food. Chem.* **2011,** *59* (16), 8806-8815.

(106) Siegel, R.; Ma, J.; Zou, Z.; Jemal, A., Cancer statistics, 2014. *CA Cancer J. Clin.* **2014,** *64* (1), 9-29.

(107) Siegel, R. L.; Miller, K. D.; Jemal, A., Cancer statistics, 2015. *CA Cancer J. Clin.* **2015,** *65* (1), 5-29.

(108) Drucker, E.; Krapfenbauer, K., Pitfalls and limitations in translation from biomarker discovery to clinical utility in predictive and personalised medicine. *EPMA J.* **2013,** *4* (1), 1-10.

(109) Polanski, M.; Anderson, N. L., A list of candidate cancer biomarkers for targeted proteomics. *Biomark. Insights* **2006,** *1*, 1-48.

(110) Serkova, N. J.; Spratlin, J. L.; Eckhardt, S. G., NMR-based metabolomics: translational application and treatment of cancer. *Curr. Opin. Mol. Ther.* **2007,** *9* (6), 572-585.

(111) Fossel, E. T.; Carr, J. M.; McDonagh, J., Detection of malignant tumors. *New Engl. J. Med.* **1986,** *315* (22), 1369-1376.

(112) Spratlin, J. L.; Serkova, N. J.; Eckhardt, S. G., Clinical applications of metabolomics in oncology: a review. *Clin. Cancer. Res.* **2009,** *15* (2), 431-440.

(113) Kim, Y. S.; Maruvada, P., Frontiers in metabolomics for cancer research: proceedings of a National Cancer Institute workshop. *Metabolomics* **2008,** *4* (2), 105-113.

(114) Issaq, H. J.; Waybright, T. J.; Veenstra, T. D., Cancer biomarker discovery: opportunities and pitfalls in analytical methods. *Electrophoresis* **2011,** *32* (9), 967-975.

(115) Florian, C. L.; Preece, N. E.; Bhakoo, K. K.; Williams, S. R.; Noble, M., Characteristic metabolic profiles revealed by ^{1}H NMR spectroscopy for three types of human brain and nervous system tumours. *NMR Biomed.* **1995,** *8* (6), 253-264.

(116) Maxwell, R. J.; Martínez-Pérez, I.; Cerdán, S.; Cabañas, M. E.; Arús, C.; Moreno, À.; Capdevila, A.; Ferrer, E.; Bartomeus, F.; Aparicio, A., Pattern recognition analysis of ^{1}H NMR spectra from perchloric acid extracts of human brain tumor biopsies. *Magn. Reson. Med.* **1998,** *39* (6), 869-877.

(117) Qiu, Y.; Cai, G.; Su, M.; Chen, T.; Zheng, X.; Xu, Y.; Ni, Y.; Zhao, A.; Xu, L. X.; Cai, S., Serum metabolite profiling of human colorectal cancer using GC-TOFMS and UPLC-QTOFMS. *J. Proteome Res.* **2009,** *8* (10), 4844-4850.

(118) Nishiumi, S.; Kobayashi, T.; Ikeda, A.; Yoshie, T.; Kibi, M.; Izumi, Y.; Okuno, T.; Hayashi, N.; Kawano, S.; Takenawa, T., A novel serum metabolomics-based diagnostic approach for colorectal cancer. *PLoS One* **2012,** *7* (7), e40459.

(119) Corominas-Faja, B.; Quirantes-Piné, R.; Oliveras-Ferraros, C.; Vazquez-Martin, A.; Cufí, S.; Martin-Castillo, B.; Micol, V.; Joven, J.; Segura-Carretero, A.; Menendez, J. A., Metabolomic fingerprint reveals that metformin impairs one-carbon metabolism in a manner similar to the antifolate class of chemotherapy drugs. *Aging* **2012,** *4* (7), 480.

(120) Glunde, K.; Serkova, N. J., Therapeutic targets and biomarkers identified in cancer choline phospholipid metabolism. *Pharmacogenomics* **2006,** *7* (7), 1109-1123.

(121) Mueller-Lisse, U. G.; Swanson, M. G.; Vigneron, D. B.; Hricak, H.; Bessette, A.; Males, R. G.; Wood, P. J.; Noworolski, S.; Nelson, S. J.; Barken, I., Time-dependent effects of hormone-deprivation therapy on prostate metabolism as detected by combined magnetic resonance imaging and 3D magnetic resonance spectroscopic imaging. *Magn. Reson. Med.* **2001,** *46* (1), 49-57.

(122) Folger, O.; Jerby, L.; Frezza, C.; Gottlieb, E.; Ruppin, E.; Shlomi, T., Predicting selective drug targets in cancer through metabolic networks. *Mol. Syst. Biol.* **2011,** *7* (1), 501.

(123) Eisner, T.; Meinwald, J., Chemical ecology. *Proc. Natl. Acad. Sci. U. S. A.* **1995,** *92* (1), 1.

(124) Prince, E. K.; Pohnert, G., Searching for signals in the noise: metabolomics in chemical ecology. *Anal. Bioanal. Chem.* **2010,** *396* (1), 193-197.

(125) Hay, M. E.; Kubanek, J., Community and ecosystem level consequences of chemical cues in the plankton. *J. Chem. Ecol.* **2002,** *28* (10), 2001-2016.

(126) Fumagalli, E.; Baldoni, E.; Abbruscato, P.; Piffanelli, P.; Genga, A.; Lamanna, R.; Consonni, R., NMR techniques coupled with multivariate statistical analysis: tools to analyse *Oryza sativa* metabolic content under stress conditions. *J. Agron. Crop. Sci.* **2009,** *195* (2), 77-88.

(127) Hirai, M. Y.; Yano, M.; Goodenowe, D. B.; Kanaya, S.; Kimura, T.; Awazuhara, M.; Arita, M.; Fujiwara, T.; Saito, K., Integration of transcriptomics and metabolomics for understanding of global responses to nutritional stresses in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U. S. A.* **2004,** *101* (27), 10205-10210.

(128) Jones, O. A.; Dondero, F.; Viarengo, A.; Griffin, J. L., Metabolic profiling of *Mytilus galloprovincialis* and its potential applications for pollution assessment. *Mar. Ecol. Prog. Ser.* **2008,** *369*, 169-179.

(129) Choi, Y.; Kim, H.; Linthorst, H.; Hollander, J.; Lefeber, A.; Erkelens, C.; Nuzillard, J.; Verpoorte, R., NMR metabolomics to revisit the tobacco mosaic virus infection in *Nicotiana tabacum* leaves. *J. Nat. Prod.* **2006,** *69* (5), 742.

(130) Schroeder, F. C.; del Campo, M. L.; Grant, J. B.; Weibel, D. B.; Smedley, S. R.; Bolton, K. L.; Meinwald, J.; Eisner, T., Pinoresinol: A lignol of plant origin serving for defense in a caterpillar. *Proc. Natl. Acad. Sci. U. S. A.* **2006,** *103* (42), 15497-15501.

(131) Peiris, D.; Dunn, W. B.; Brown, M.; Kell, D. B.; Roy, I.; Hedger, J. N., Metabolite profiles of interacting mycelial fronts differ for pairings of the wood decay basidiomycete fungus, *Stereum hirsutum* with its competitors *Coprinus micaceus* and *Coprinus disseminatus*. *Metabolomics* **2008,** *4* (1), 52-62.

(132) Broadhurst, D. I.; Kell, D. B., Statistical strategies for avoiding false discoveries in metabolomics and related experiments. *Metabolomics* **2006,** *2* (4), 171-196.

(133) Dunn, W. B.; Wilson, I. D.; Nicholls, A. W.; Broadhurst, D., The importance of experimental design and QC samples in large-scale and MS-driven untargeted metabolomic studies of humans. *Bioanalysis* **2012,** *4* (18), 2249-2264.

(134) Zelena, E.; Dunn, W. B.; Broadhurst, D.; Francis-McIntyre, S.; Carroll, K. M.; Begley, P.; O'Hagan, S.; Knowles, J. D.; Halsall, A.; Wilson, I. D., Development of a robust and repeatable UPLC−MS method for the long-term metabolomic study of human serum. *Anal. Chem.* **2009,** *81* (4), 1357-1364.

(135) Kochhar, S.; Jacobs, D. M.; Ramadan, Z.; Berruex, F.; Fuerholz, A.; Fay, L. B., Probing gender-specific metabolism differences in humans by nuclear magnetic resonance-based metabonomics. *Anal. Biochem.* **2006,** *352* (2), 274-281.

(136) Slupsky, C. M.; Rankin, K. N.; Wagner, J.; Fu, H.; Chang, D.; Weljie, A. M.; Saude, E. J.; Lix, B.; Adamko, D. J.; Shah, S., Investigations of the effects of gender, diurnal variation, and age in human urinary metabolomic profiles. *Anal. Chem.* **2007,** *79* (18), 6995-7004.

(137) Carraro, S.; Rezzi, S.; Reniero, F.; Héberger, K.; Giordano, G.; Zanconato, S.; Guillou, C.; Baraldi, E., Metabolomics applied to exhaled breath condensate in childhood asthma. *Am. J. Respir. Crit. Care Med.* **2007,** *175* (10), 986-990.

(138) Tiziani, S.; Emwas, A.-H.; Lodi, A.; Ludwig, C.; Bunce, C. M.; Viant, M. R.; Günther, U. L., Optimized metabolite extraction from blood serum for ^{1}H nuclear magnetic resonance spectroscopy. *Anal. Biochem.* **2008,** *377* (1), 16-23.

(139) Want, E. J.; O'Maille, G.; Smith, C. A.; Brandon, T. R.; Uritboonthai, W.; Qin, C.; Trauger, S. A.; Siuzdak, G., Solvent-dependent metabolite distribution, clustering, and protein extraction for serum profiling with mass spectrometry. *Anal. Chem.* **2006,** *78* (3), 743-752.

(140) Dietmair, S.; Timmins, N. E.; Gray, P. P.; Nielsen, L. K.; Krömer, J. O., Towards quantitative metabolomics of mammalian cells: development of a metabolite extraction protocol. *Anal. Biochem.* **2010,** *404* (2), 155-164.

(141) Masson, P.; Alves, A. C.; Ebbels, T. M.; Nicholson, J. K.; Want, E. J., Optimization and evaluation of metabolite extraction protocols for untargeted metabolic profiling of liver samples by UPLC-MS. *Anal. Chem.* **2010,** *82* (18), 7779-7786.

(142) Xia, J.; Broadhurst, D. I.; Wilson, M.; Wishart, D. S., Translational biomarker discovery in clinical metabolomics: an introductory tutorial. *Metabolomics* **2013,** *9* (2), 280-299.

(143) Pluskal, T.; Castillo, S.; Villar-Briones, A.; Orešič, M., MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* **2010,** *11*, 395.

(144) Gowda, H.; Ivanisevic, J.; Johnson, C. H.; Kurczy, M. E.; Benton, H. P.; Rinehart, D.; Nguyen, T.; Ray, J.; Kuehl, J.; Arevalo, B.; Westenskow, P. D.; Wang, J.; Arkin, A. P.; Deutschbauer, A. M.; Patti, G. J.; Siuzdak, G., Interactive XCMS online: simplifying advanced metabolomic data processing and subsequent statistical analyses. *Anal. Chem.* **2014,** *86* (14), 6931-6939.

(145) Katajamaa, M.; Orešič, M., Data processing for mass spectrometry-based metabolomics. *J. Chromatogr. A* **2007,** *1158* (1), 318-328.

(146) Katajamaa, M.; Miettinen, J.; Orešič, M., MZmine: toolbox for processing and visualization of mass spectrometry based molecular profile data. *Bioinformatics* **2006,** *22* (5), 634-636.

(147) Boccard, J.; Veuthey, J. L.; Rudaz, S., Knowledge discovery in metabolomics: an overview of MS data handling. *J. Sep. Sci.* **2010,** *33* (3), 290-304.

(148) van den Berg, R. A.; Hoefsloot, H. C.; Westerhuis, J. A.; Smilde, A. K.; van der Werf, M. J., Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics* **2006,** *7* (1), 142.

(149) Parsons, H. M.; Ludwig, C.; Günther, U. L.; Viant, M. R., Improved classification accuracy in 1-and 2-dimensional NMR metabolomics data using the variance stabilising generalised logarithm transformation. *BMC Bioinformatics* **2007,** *8*, 234.

(150) Tsugawa, H.; Tsujimoto, Y.; Arita, M.; Bamba, T.; Fukusaki, E., GC/MS based metabolomics: development of a data mining system for metabolite identification by using soft independent modeling of class analogy (SIMCA). *BMC Bioinformatics* **2011,** *12*, 131.

(151) Mahadevan, S.; Shah, S. L.; Marrie, T. J.; Slupsky, C. M., Analysis of metabolomic data using support vector machines. *Anal. Chem.* **2008,** *80* (19), 7562-7570.

(152) Trygg, J.; Holmes, E.; Lundstedt, T., Chemometrics in metabonomics. *J. Proteome Res.* **2007,** *6* (2), 469-479.

(153) Guan, W.; Zhou, M.; Hampton, C. Y.; Benigno, B. B.; Walker, L. D.; Gray, A.; McDonald, J. F.; Fernández, F. M., Ovarian cancer detection from metabolomic liquid chromatography/mass spectrometry data by support vector machines. *BMC Bioinformatics* **2009,** *10*, 259.

(154) Rajalahti, T.; Arneberg, R.; Berven, F. S.; Myhr, K.-M.; Ulvik, R. J.; Kvalheim, O. M., Biomarker discovery in mass spectral profiles by means of selectivity ratio plot. *Chemometrics Intellig. Lab. Syst.* **2009,** *95* (1), 35-48.

(155) Wiklund, S.; Johansson, E.; Sjöström, L.; Mellerowicz, E. J.; Edlund, U.; Shockcor, J. P.; Gottfries, J.; Moritz, T.; Trygg, J., Visualization of GC/TOF-MS-based metabolomics data for identification of biochemically interesting compounds using OPLS class models. *Anal. Chem.* **2008,** *80* (1), 115-122.

(156) Brown, M.; Dunn, W. B.; Dobson, P.; Patel, Y.; Winder, C.; Francis-McIntyre, S.; Begley, P.; Carroll, K.; Broadhurst, D.; Tseng, A., Mass spectrometry tools and metabolite-specific databases for molecular identification in metabolomics. *Analyst* **2009,** *134* (7), 1322-1332.

(157) Patti, G. J.; Yanes, O.; Siuzdak, G., Innovation: Metabolomics: the apogee of the omics trilogy. *Nat. Rev. Mol. Cell Biol.* **2012,** *13* (4), 263-269.

(158) Douglas, D.; Konenkov, N., Ion source emittance influence on the transmission of a quadrupole operated in the second stability region. *J. Am. Soc. Mass. Spectrom.* **1998,** *9* (10), 1074-1080.

(159) Horai, H.; Arita, M.; Kanaya, S.; Nihei, Y.; Ikeda, T.; Suwa, K.; Ojima, Y.; Tanaka, K.; Tanaka, S.; Aoshima, K., MassBank: a public repository for sharing mass spectral data for life sciences. *J. Mass Spectrom.* **2010,** *45* (7), 703-714.

PART I: APPLICATIONS OF MASS SPECTROMETRY-BASED UNTARGETED METABOLOMICS

CHAPTER 2. ULTRA PERFORMANCE LIQUID CHROMATOGRAPHY−MASS SPECTROMETRY SERUM METABOLOMICS DETECTION OF PROSTATE CANCER

Reprinted with permission from
Zang, X.‡; Jones, C. M.‡; Long, T. Q.; Monge, M. E.; Zhou, M.; Walker, L. D.; Mezencev, R.; Gray, A.; McDonald, J. F.; Fernández, F. M., Feasibility of Detecting Prostate Cancer by Ultra Performance Liquid Chromatography−Mass Spectrometry Serum Metabolomics. *J. Proteome Res.* **2014**, *13* (17), 3444-3454. Copyright © 2014 American Chemical Society.
‡equal contributing author

This chapter describes research conducted by multiple persons. C. M. Jones and M. Zhou optimized the sample preparation and UPLC-MS analysis methods in addition to acquiring the UPLC-MS data. C. M. Jones processed the UPLC-MS data. C. M. Jones, X. Zang, and M. E. Monge tentatively identified metabolites and conducted UPLC-MS/MS experiments. X. Zang confirmed all metabolite identifications, conducted chemical standard validation experiments with assistance from M. E. Monge, and determined the biological functions of the identified metabolites with assistance from R. Mezencev. T. Q. Long performed all multivariate analyses.

2.1 Abstract

Prostate cancer (PCa) is the second leading cause of cancer-related mortality in men. The prevalent diagnosis method is based on the serum Prostate-Specific Antigen (PSA) screening test, which suffers from low specificity, over-diagnosis and over-treatment. In this work, untargeted metabolomic fingerprinting of age-matched serum samples from prostate cancer patients and healthy individuals was performed using ultra performance liquid chromatography coupled to high resolution tandem mass spectrometry (UPLC-MS/MS) and machine learning methods. A metabolite-based *in vitro* diagnostic multivariate index assay (IVDMIA) was developed to predict the presence of PCa in serum samples with high classification sensitivity, specificity and accuracy. A panel of 40 metabolic spectral features was found to be differential with

92.1% sensitivity, 94.3% specificity, and 93.0% accuracy. The performance of the IVDMIA was higher than the prevalent PSA test. Within the discriminant panel, 31 metabolites were identified by MS and MS/MS, with 10 further confirmed chromatographically by standards. Numerous discriminant metabolites were mapped in the steroid hormone biosynthesis pathway. The identification of fatty acids, amino acids, lysophospholipids, and bile acids provided further insights into the metabolic alterations associated with the disease. With additional work, the results presented here show potential towards implementation in clinical settings.

2.2 Prostate Cancer Detection

2.2.1 Current Diagnostic Methodology

Prostate cancer (PCa) is the 2nd leading cause of cancer-related mortality in men worldwide, with 30,000 deaths per year in the U.S. alone.[1] The prevalent diagnosis method is based on the triad of digital rectal examination, blood Prostate-Specific Antigen (PSA) measurement, and ultrasound-guided prostate biopsy. Although the introduction of PSA screening decreased mortality by 4% between 1994 and 2006,[2] the use of PSA as a diagnostic serum marker still presents several drawbacks. The concentration of this protein in the blood stream increases during the development of cancer, but also can be secreted as a result of benign prostatic hyperplasia, prostatitis, or other traumas to prostate cells.[3] Therefore, this method suffers from low specificity and consequent over-diagnosis and over-treatment.[4-7] Moreover, approximately 15% of patients with PCa have PSA values lower than the commonly used cutoff point of 4.0 ng mL^{-1}, leaving many cases undetected.[8]

2.2.2 Overview of Metabolic Prostate Cancer Detection

The shortcomings of PSA as a diagnostic PCa serum biomarker have led to an increased interest in using untargeted metabolomic fingerprinting/profiling to discover new differential metabolic biomarkers that could improve the specificity of PCa diagnosis.[9] Metabolic biomarkers are used as a routine tool in screening newborns for the presence of inborn errors of metabolism by means of tandem mass spectrometry;[10, 11] however, global metabolite profiling of PCa patients still remains at an early stage, and there is no biomarker panel currently in use for clinical testing.[9] Current research has shown some evidence of metabolic alterations associated with PCa. Tissue sarcosine levels have been suggested as a potential biomarker for the aggressive form of the disease in a metabolomic profiling study using both liquid and gas chromatography coupled to mass spectrometry (LC-MS and GC-MS).[12] Its concentration in prostate cancer-related tissue specimens was highly increased during PCa progression to metastasis, but differences in urine were much less marked.[12] These results have been very prominent but somewhat controversial as other targeted studies failed in the attempt of differentiating healthy individuals from cancer patients based on sarcosine concentration in biological fluids and cancerous tissues.[9, 13-15] The analysis of cancerous tissues by proton high-resolution magic angle spinning nuclear magnetic resonance (NMR) spectroscopy has shown a decrease in the concentrations of citrate and polyamines, and increases in cholines, glycerophospholipids, and lactate concentrations during PCa proliferation.[16, 17] Increased levels of cholesterol as well as alterations in amino acid metabolism were detected in metastatic bone samples by GC-MS.[18] However, all of these studies included too few patients to offer strong leads on the metabolic alterations

associated with PCa. A panel of plasma lipids that included phosphatidylethanolamines, ether-linked phosphatidylethanolamines, and ether-linked phosphatidylcholines was proposed to discriminate PCa patients from healthy groups through direct infusion electrospray ionization tandem MS.[19] The authors demonstrated that a combination of multiple biomarkers with multivariate analysis and various classification algorithms yielded better predictive power for the diagnosis of PCa than univariate analysis of single lipid species. However, the predictive power was not compared with that of PSA, as this information was not available at the time of cohort design.[19] More robust metabolic models still need to be developed for improved understanding of disease progression, and more reliable PCa detection.

2.3 Experimental Details

2.3.1 Chemicals

Healthy human blood serum (S7023-50 mL) and acetic acid (≥ 99.7%) were purchased from Sigma-Aldrich Corp. (St. Louis, MO, USA). Omnisolv LC-MS grade acetonitrile, Omnisolv high purity dichloromethane and HPLC grade acetone were purchased from EMD (Billerica, MA, USA). LC-MS grade methanol and 2-propanol were purchased from J.T. Baker Avantor Performance Materials, Inc. (Center Valley, PA, USA). Ultrapure water with 18.2 MΩ cm resistivity (Barnstead Nanopure UV ultrapure water system, USA) was used to prepare mobile phases. Uric acid (≥99%), azelaic acid (98%), undecanedioic acid (97%), heptadecanoic acid (≥98%) and decanoic acid (≥98%) were purchased from Sigma-Aldrich Corp. (St. Louis, MO, USA). Hexadecanedioic acid (98%) was purchased from Ark Pharm, Inc. (Libertyville, IL, USA). Phenylalanyl

phenylalanine was purchased from MP Biomedicals (Solon, OH, USA). Phenylacetyl glutamine was purchased from Bachem (Hauptstrasse, Bubendorf, Sitzerland). Indoxyl sulfate potassium was purchased from Alfa Aesar (Ward Hill, MA, USA). 1-stearoyl-2-hydroxy-sn-glycero-3-phosphocholine/ lysoPC (18:0/0:0) was purchased from Avanti Polar Lipids, Inc. (Alabaster, AL, USA).

2.3.2 Patient Cohort Description

Age-matched blood serum samples were obtained from 64 PCa patients (age range 49-65, mean age 59 ± 4 years) and 50 healthy individuals (age range: 45-76, mean age 57 ± 7 years). At the 0.05 level, the population means were not significantly different with the two-sample *t*-test. The cohort ethnicity was as follows: 28 African American (24.6%); 76 Caucasian (66.7%); 5 Hispanic (4.4%); 2 Asian (1.8%); 2 Jewish ancestry (1.8%); and 1 unknown (0.9%). After approval by the Institutional Review Board (IRB), blood samples were collected at Saint Joseph´s Hospital of Atlanta (GA, USA) by venipuncture from each donor into evacuated blood collection tubes that contained no anticoagulant. Serum was obtained by centrifugation at 5000 rpm for 5 min at 4 °C. Immediately after centrifugation, 200 µL aliquots of serum were frozen and stored at -80 °C for further use. The sample collection and storage procedures for PCa patients and healthy individuals were identical. Gleason scores based on the microscopic glandular patterns of biopsy specimens were obtained for 61 PCa patients.

2.3.3 Sample Preparation and Ultra Performance Liquid Chromatography-Mass Spectrometry Protocols

A stock sample of healthy human blood serum was used to develop the serum sample preparation protocol and ultra performance liquid chromatography-mass spectrometry (UPLC-MS) method. Serum samples were thawed on ice, and protein precipitation was performed by the addition of a mixture of acetone, acetonitrile and methanol (1:1:1 v/v) to 100 µL of serum in a 3:1 volume ratio. Samples were vortex-mixed for 20 s, and centrifuged at 16000 × *g* for 5 min. After centrifugation, 800 µL of dichloromethane were added to 350 µL of supernatant, and vortex-mixed. Following the addition of 250 µL of deionized water, samples were vortex-mixed again to extract the non-polar lipid fraction. The aqueous phase was used for metabolite analysis by UPLC-MS. Samples were randomly separated into 7 batches and consecutively analyzed. The instrument was calibrated before analysis and solvent and sample preparation blanks were jointly analyzed with the samples in a random order.

UPLC-MS analysis was performed using a Waters ACQUITY Ultra Performance LC (Waters Corporation, Manchester, UK) system, fitted with a Waters ACQUITY UPLC BEH C_{18} column (2.1 × 50 mm, 1.7 µm particle size), and coupled to a high-resolution accurate mass (HRAM) Synapt G2 High Definition Mass Spectrometry (HDMS) system (Waters Corporation, Manchester, UK). The Synapt G2 HDMS is a hybrid quadrupole-ion mobility-orthogonal acceleration time-of-flight instrument with typical resolving power of 20,000 FWHM M/Δ*m* and mass accuracy of 9 ppm at m/*z* 554.2615. The instrument was operated in negative ion mode with a probe capillary voltage of 2.3 kV, and a sampling cone voltage of 45 V. The source and desolvation

temperatures were 120 °C and 350 °C, respectively; and the nitrogen desolvation flow rate was 650 L h^{-1}. The mass spectrometer was calibrated across the range of m/z 50-1800 using a 0.5 mM sodium formate solution prepared in 90:10 2-propanol:water v/v. Data were mass corrected during acquisition using a leucine enkephalin reference spray (LockSpray) infused at 2 µL min^{-1}. Data were acquired in the 50-1750 m/z range and the scan time was set to 1 s. Data acquisition and processing was carried out using MassLynx v4.1. The chromatographic method for sample analysis involved elution with acetonitrile (mobile phase A) and water with 0.1% acetic acid (mobile phase B) using the following gradient program: 0-1 min 0-10% A; 1-2.5 min 10-15% A; 2.5-4 min 15-22% A; 4-6 min 22-38% A; 6-9 min 38-65% A; 9-12 min 65-80% A; 12-16 min 80-100% A; 16-18 min 100% A. The flow rate was constant at 0.25 mL min^{-1} for 12 min. It was increased to 0.30 mL min^{-1} between 12 and 16 min, and from 0.30 to 0.45 mL min^{-1} between 16 and 18 min. The gradient was returned to its initial conditions over a period of 8 minutes after each sample injection. The column temperature was set to 35 °C, the autosampler tray temperature was set to 5 °C, and the injection volume was 10 µL. UPLC-MS/MS experiments were performed by acquiring mass spectra with applied voltages between 5 and 50 V in the trap cell, using ultra purity argon (≥ 99.999%) as the collision gas.

2.3.4 Data Analysis

After UPLC-MS analysis, metabolic features (retention time (R_t), m/z pairs) were extracted from chromatograms using MarkerLynx XS software. This procedure involved chromatogram alignment, peak picking and integration, peak area extraction, and normalization. The matrix containing sample peak areas for each feature (R_t, m/z) was utilized to build a model for sample classification and to find the minimum set of

discriminant features by means of linear support vector machines (SVMs).[20] This supervised classification technique is effective at handling high dimensionality data as those produced in the present work. For a binary classification problem, linearly-separable samples represented as a row vector $\boldsymbol{x}$, had membership of two classes g (= H or D), where H stands for healthy and D for PCa disease with labels c = -1 for class H, and +1 for class D. To build the classification model, 70% of the samples were randomly selected as a training set, and 30% as a test set. Within the training set, 10% of samples were used for validation and to find the minimum set of discriminant features that maximized accuracy in the classification through a recursive feature elimination (RFE) method.[21] The decision function that separated the two classes, defined here as the IVDMIA "PCa metabolic score", was as follows:

$$PCa\ metabolic\ score = b + \sum_{j=1}^{J} w_j x_{ij} \quad [1]$$

$$g(\boldsymbol{x}_i) = sgn(\boldsymbol{w}\boldsymbol{x}_i' + b) = sgn(PCa\ metabolic\ score) \quad [2]$$

where w and b are the weight and bias parameters that were determined from the training set and J is the total number of features. The sign of the PCa metabolic score determined which class a sample was assigned to: class H if negative and class D if positive. In this classification function, the two classes were divided in the dataspace by a hyperplane $\boldsymbol{w}\boldsymbol{x}' + b = 0$ that maximized the margins between samples of different classes. The margin between the two classes was defined such that:

$$\boldsymbol{w}\boldsymbol{x}' + b \geq 1, \qquad c = +1 \quad [3]$$

$$\boldsymbol{w}\boldsymbol{x}' + b \leq -1, \qquad c = -1 \quad [4]$$

To estimate the classification and feature selection performance, ten iterative validations were performed to randomly select the training and test sets. The statistical significance of the model was further assessed through hypothesis testing by permutation tests. No assumptions were made in this non-parametric approach to hypothesis testing regarding the data distribution, and the *p*-value was computed as the cumulative sum using the empirical distribution. Two permutation tests were performed using 100 permutation samples with the following null hypothesis:

i) Null hypothesis 1: feature and labels (positive/negative) are independent (i.e. indifference when class labels are permutated).

ii) Null hypothesis 2: features are independent within each class (i.e. indifference when value of each features are permutated within each class).

If the *p*-value < α (α = 0.05), the null hypothesis H_0 was rejected; otherwise the observed result was not statistically significant.

Additionally, Principal Component Analysis (PCA) was used to evaluate the performance of all extracted metabolic features or subsets of them in an unsupervised manner with MATLAB R2011b (Version 7.13.0, The MathWorks, Inc., Natick, MA, USA) and the PLS Toolbox (v.6.71, Eigenvector Research, Inc., Wenatchee, WA, USA). Data were preprocessed by autoscaling.

2.3.5 Discriminant Metabolite Identification Procedure

Compound identification was attempted for the 40 discriminant features remaining after the feature selection processes. Due to the biological complexity of serum samples, adduct ion analysis was first performed to ensure the unambiguous assignment

of the signal of interest in each mass spectrum. Adduct ions corresponding to SVM-selected variables that were investigated in the mass spectra included $[M - H]^-$, $[M + Cl]^-$,$[M + Br]^-$, $[M + CH_3COO]^-$, $[M + HCOO]^-$, $[M + CF_3COO]^-$ $[M + Na - 2H]^-$, $[M + K - 2H]^-$, $[M - H_2O - H]$, $[M + H_2O - H]^-$ species, which are typically observed in negative electrospray ionization mode. The expected m/z values for common adduct species were calculated and compared with the experimental values from peaks within the spectra. For spectra in which no confirmatory adducts were present, the accurate mass of the candidate neutral molecule was calculated based on the assumption that the peak of interest corresponded to $[M - H]^-$. Elemental formulae were generated based on the mass accuracy of the peak of interest and isotopic patterns with a mass error of 8 mDa, using MassLynx 4.1. The list of elements included in the search was C, H, N, O, P, S, Cl, and Br. The list of generated elemental formulae were searched against the METLIN database,[22] HMDB,[23] and MassBank[24] to determine the possible endogenous metabolite candidates. The METLIN MS/MS database and a literature survey were subsequently used to confirm the identity of putative candidates. Fragmentation patterns were also manually analyzed to discriminate between different isobaric species.

2.4 Prostate Cancer Detection Performance of the *In Vitro* Diagnostic Multivariate Assay

UPLC-MS analysis in negative ion mode allowed the interrogation of highly complex serum samples from PCa patients and healthy individuals, revealing a total of 480 features (R_t, m/z pairs). The extracted features were used to build a discriminant SVM model for sample classification. An optimum set of 51 discriminant features were

found to maximize classification accuracy through a RFE method,[21] as illustrated in Figure 2.1.

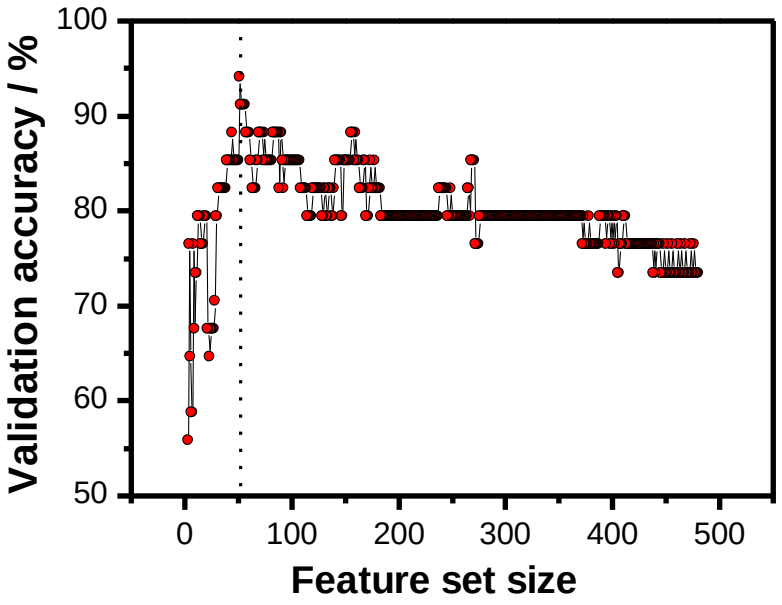

Figure 2.1: Evolution of classification accuracy for a validation sample subset consisting of 10% of the training samples as a function of the number of features retained. The minimum discriminant feature set that maximizes classification accuracy is highlighted with a dashed line.

Out of the 51 selected features, 7 were found to be only present in less than 2% of the samples; 2 features were identified as acetaminophen and its sulfite adduct, and 2 additional features were identified as adducts or fragments of other features in the subset, and were thus removed from further consideration. The optimum panel that best discriminated PCa patients from healthy individuals was thus reduced to 40 features, demonstrating that the feature selection process accomplished a high reduction in problem dimensionality.

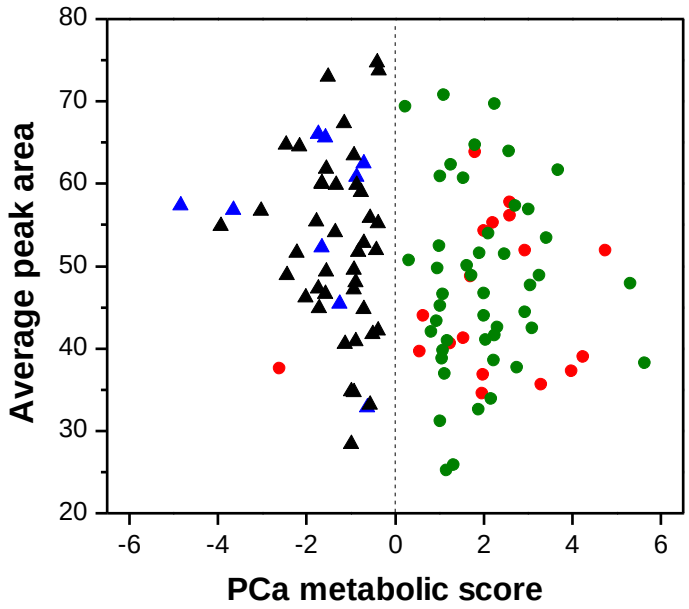

Figure 2.2: Visualization of the PCa metabolic scores obtained by SVMs in one out of 10 iterative model validations based on 40 discriminant features. Green circles correspond to PCa patients in the training set, black triangles correspond to controls in the training set, red circles correspond to PCa patients in the test set built for the iteration shown, and blue triangles correspond to healthy individuals in the test set. The dotted line shows the projection of the separating hyperplane: $\boldsymbol{wx}' + b = 0$.

Figure 2.2 illustrates the "PCa metabolic scores" obtained for the training and the test sets of randomly selected samples that were used to construct and evaluate the classification model, respectively. The separation of the two sample classes (H or D) was determined in the data space by the optimal separating hyperplane for which the margin between the most similar samples in each group was largest, illustrated with a dotted line in the figure.

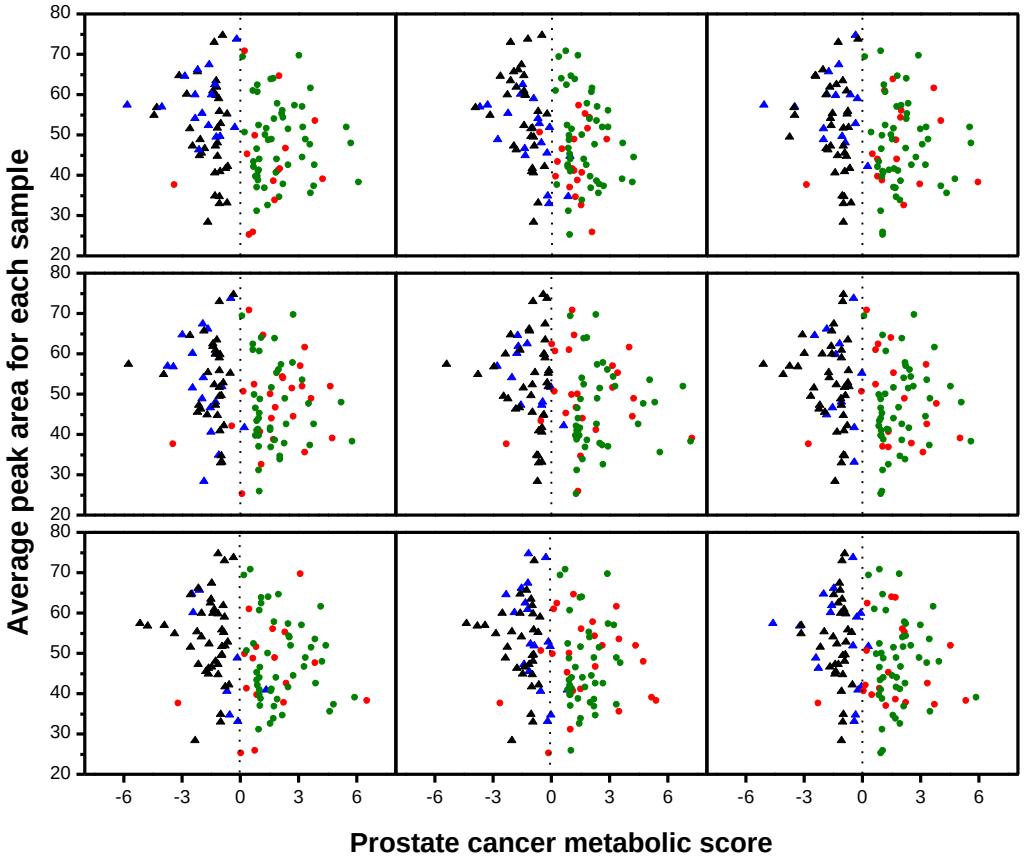

Figure 2.3: Visualization of the PCa metabolic scores obtained by SVMs in 9 out of 10 model validation iterations. Green circles correspond to PCa patients in the training set, black triangles correspond to controls in the training set, red circles correspond to PCa patients in the test set, and blue triangles correspond to healthy individuals in the test set. The dotted line shows the projection of the separating hyperplane: $\boldsymbol{wx}' + b = 0$.

The samples with scores equal to 1 or -1 are the support vectors of the model. For the particular cross-validation iteration illustrated in Figure 2.2 only one sample was misclassified as a false negative. The remaining 9 iterative validations with their respective training and test sets are illustrated in Figure 2.3. Based on these 40 discriminant features, serum samples were successfully classified as cancerous or healthy with 93.0% accuracy, 92.1% sensitivity, and 94.3% specificity. These values were

calculated as the averages from 10 distinct test sets, illustrated in Figures 2.2 and 2.3. In addition, the statistical significance of the model was further evaluated through hypothesis testing and, at 0.05 significance level, the null hypothesis was rejected for all permutations generated (p-value = 0.0099). Unambiguously, the classifier did not yield a better leave-one-out cross-validation (LOOCV) accuracy rate than the original data. These results suggest a promising approach that could form the basis for a PCa IVDMIA. In particular, of the 40 differential features, 24 were found to increase in sera from PCa patients, and 16 were found to decrease in PCa, as illustrated in Figure 2.4. It is important to underline, however, that the strength of this IVDMIA resides in the combination of multiple metabolic features using an interpretation function to yield a single, patient-specific result to be used in the disease diagnosis, and not on the average fold change of each differential feature.

To evaluate the possible risk of data overfitting by SVMs,[20] a simple unsupervised approach was also used to examine the dataset. PCA score plots were generated for both the 40 discriminant features set obtained by SVMs and the starting set of 480 features. Figure 2.5 shows the results for each case.

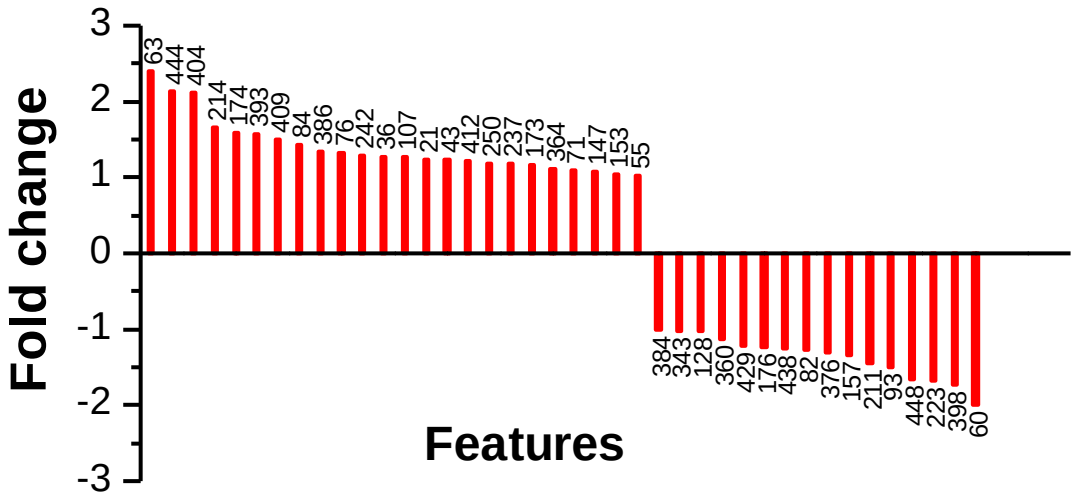

Figure 2.4: Fold change of average peak areas of each discriminant feature. Positive fold changes are calculated as the ratio of average peak areas between PCa patients and healthy individuals, and negative fold changes are calculated as the negative ratio of average peak areas between healthy individuals and PCa patients. Features are labeled with their codes.

Using the best 40 features, three principal components containing 33.6% of the total variance provided a good degree of separation between classes, as illustrated in Figure 2.5a. Sample separation in the PCA score plot was mainly achieved by the contribution of PC3. Loadings for PC3 are displayed in Figure 2.5b. Interestingly, Figure 2.5c shows that PCA does not provide any distinguishable clustering when applied to the initial set of 480 features, further supporting the use of RFE and SVMs when handling high dimensionality data volumes as those in the present work. Given the clustering observed in PCA when using the 40 discriminant feature subset, the risk of the high classification accuracy of SVM models being a product of overfitting is greatly diminished.

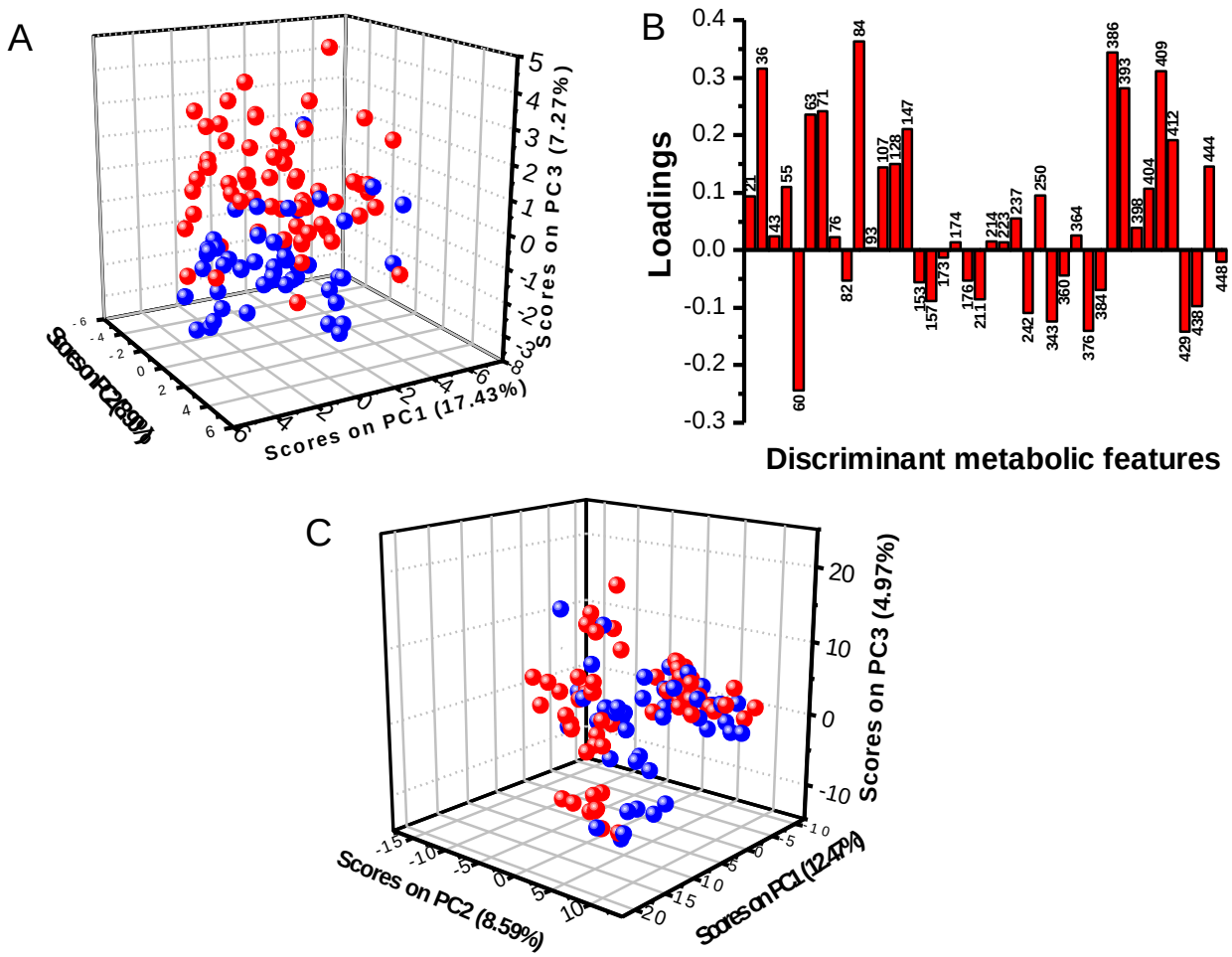

Figure 2.5: Principal Component Analysis (PCA). A: PCA score plot using only the 40 discriminant features obtained by RFE-SVMs. B: Loadings plot obtained for PC3 using the 40 discriminant features, each labeled with their code. C: PCA score plot of the initial set of 480 features. Samples from PCa patients are illustrated with red circles and samples from healthy individuals are illustrated with blue circles.

2.5 *In Vitro* Diagnostic Multivariate Assay Versus Prostate Specific Antigen Diagnosis

The Gleason scores for the PCa patients, summarized in Table 2.1, indicate that the most common tumor patterns presented by the patients derived from moderate to aggressive cancers.

Table 2.1: Gleason scores for PCa patients.

Gleason Sum	# of patients; (%)
3+3 = 6	13; (20.3)
3+4 = 7	27; (42.2)
3+3 = 6; tert=4	6; (9.4)
4+3 = 7	3; (4.7)
3+4=7; tert=5	2; (3.1)
4+3=7; tert= 5	2; (3.1)
(R) 3+4 = 7; (L) 4+3 = 7	1; (1.6)
4+5 = 9	1; (1.6)
5+4 = 9	1; (1.6)
(R)3+4=7; tert= 5; (L) 3+3=6	1; (1.6)
(R)3+4=7; (L) 4+3 = 7; tert= 5	1; (1.6)
3+5=8; tert= 4	1; (1.6)
(R) 3+3 = 6; (L) 4+5 = 9	1; (1.6)
(R)3+3=6; (L)3+3=6 tert= 4	1; (1.6)

However, the PSA test performed at surgery did not follow this histological evidence for the entire PCa cohort, as 33% of patients with PCa (n = 20) had PSA values lower than the commonly used cutoff point of 4.0 ng mL^{-1}. Figure 2.6 compares PSA and IVDMIA results in terms of true positive and false negative outputs, highlighted in red and black, respectively. The IVDMIA outputs provided by the randomly-selected 10 test sets are visualized as box plots in the figure, and show that the IVDMIA was able to correctly predict 100% of the true positives that were incorrectly diagnosed as negatives by the PSA test. The false negative results provided by the IVDMIA derived from one sample that was misclassified in all test sets and 4 samples that were misclassified in at least one test set. The classification performance obtained with this cohort shows promise towards prostate cancers that would go undetected by the PSA method. The use of multiple discriminant features by this metabolic IVDMIA yields higher predictive power for PCa diagnosis than the univariate analysis of a single marker such as with the PSA method.

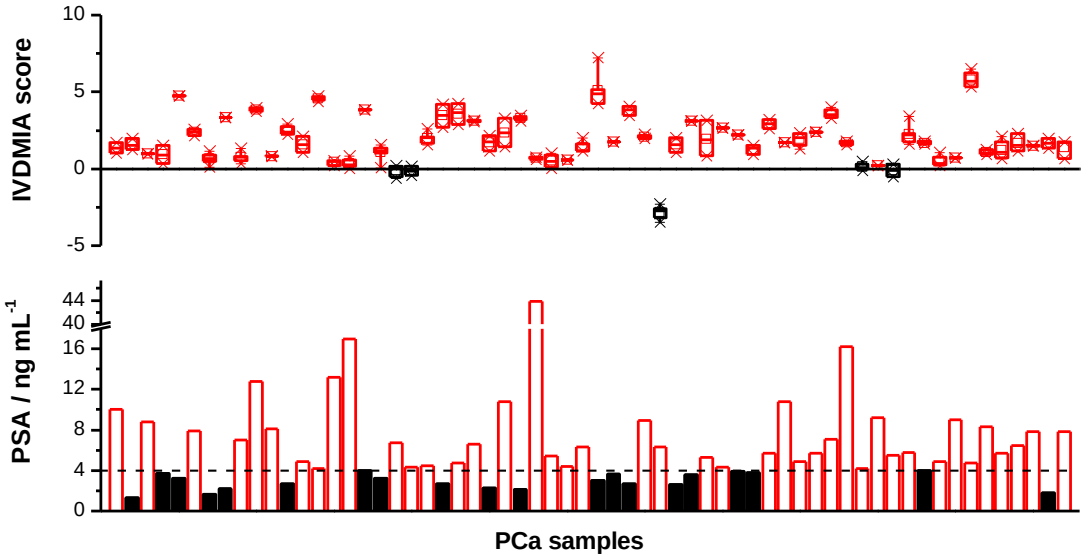

Figure 2.6: Comparison of IVDMIA vs. PSA diagnosis performance for 62 PCa patients. True positive and false negative outputs are highlighted in red and black, respectively. The cutoff point of 4.0 ng mL^{-1} used in PSA-based diagnosis is indicated with a dotted line. The IVDMIA score output is presented as box plots in the figure, each of which is generated by results obtained for each of the 10 test sets where each sample was selected for validation. No comparison is shown for 2 of the 64 PCa samples as they were not randomly selected in any of the 10 cross-validation iterations.

2.6 *In Vitro* Diagnostic Multivariate Assay Potential in Clinical Applications

To determine the fraction of samples in which the discriminant features were present, and to evaluate the feasibility of implementing the PCa IVDMIA in clinical laboratory settings through targeted triple-quadrupole mass spectrometry-based assays, smaller subgroups of the optimum 40 discriminant features, subsequently referred to as "panel A", were investigated (Table 2.2).

Table 2.2: Discriminant feature (sub)panels for PCa detection.

Panel	Accuracy	Specificity	Sensitivity	Discriminant features (#)	Discriminant Feature Codes	% Healthy samples, % PCa samples
A	93.0	94.3	92.1	40	147, 36, 71, 211, 60, 55, 107, 409, 250, 223, 386, 438, 157, 63, 176, 82, 393, 173, 84, 412, 43, 376, 343, 429, 384, 76, 444, 214, 128, 93, 398, 360, 448, 174, 153, 21, 364, 404, 242, 237	>0%; >0%
B	91.2	90.6	91. 7	38	147, 36, 71, 211, 60, 55, 107, 409, 250, 223, 386, 438, 157, 63, 176, 82, 393, 173, 84, 412, 43, 376, 343, 429, 384, 76, 444, 214, 128, 93, 398, 360, 448, 174, 153, 21, 364, 404	>50%; >0%
C	90.2	87.2	91.8	35	147, 36, 71, 211, 60, 55, 107, 409, 250, 223, 386, 438, 157, 63, 176, 82, 393, 173, 84, 412, 43, 376, 343, 429, 384, 76, 444, 214, 128, 93, 398, 360, 448, 174, 153	>50%; >50% and >0%; >50%
D	86.1	87.2	85.3	28	147, 36, 71, 211, 60, 55, 107, 409, 250, 223, 386, 438, 157, 63, 176, 82, 393, 173, 84, 412, 43, 376, 343, 429, 384, 76, 444, 214	>0%; >70%
E	84.4	80.0	85.8	25	147, 36, 71, 211, 60, 55, 107, 409, 250, 223, 386, 438, 157, 63, 176, 82, 393, 173, 84, 412, 43, 376, 343, 429, 384	>70%; >70% and >70%; >0%
F	85.0	80.0	88.8	22	147, 36, 71, 60, 55, 409, 223, 386, 438, 157, 63, 176, 82, 393, 173, 84, 412, 43, 376, 343, 429, 384	>90%; >0%
G	80.0	81.0	79.3	17	147, 36, 71, 60, 55, 409, 386, 438, 157, 176, 82, 393, 173, 84, 343, 429, 384	>90%; >90% and >0%; >90%

These subpanels were chosen to provide the minimum number of features that collectively captured metabolic PCa patterns with a high level of accuracy, specificity and sensitivity. The selection of these additional subpanels was based on the fraction of features that were present in 50, 70 or 90% of the entire sample cohort, in either PCa patients or healthy controls. Table 2.2 summarizes the different panels constructed following these criteria, with their corresponding subset of discriminant features. These panels were used to build new SVM models, and cross-validated to provide average values of accuracy, specificity and sensitivity from 10-independent randomly-selected training and testing sets. Thirty eight out of 40 discriminant features were present in more than 50% of healthy controls (Panel B) and 35 out of 40 were present in more than 50% of PCa samples (Panel C), providing similar accuracy, specificity, and sensitivity as panel A. When the criterion for feature presence was made more stringent, from panel A to panel G; the accuracy, specificity and sensitivity decreased by only ~10%, suggesting the robust biological role that the detected features might have. In other words, the different feature subpanels were not highly sensitive to a reduction in the number of discriminant features, suggesting that the smaller number of metabolites contained in subpanel G could still be potentially used to build a more focused, simpler IVDMIA for PCa detection in a clinical setting. To further test this finding, another SVM model was created with only those 13 features that could be confidently assigned to metabolites in subpanel G by HRAM MS and MS/MS (Table 2.3). It was found that this model still provided high classification sensitivity (88.3%), specificity (80.3%), and accuracy (85.0%). The mass spectrometric assay for a model of this type would be much simpler to implement in a targeted fashion due to the reduced number of transitions that a UPLC-

MS/MS triple quadrupole method would require, allowing higher analysis throughput and minimizing cost.

The set of 40 SVM weights obtained for panel A from the optimal classification model are shown in Figure 2.7. The figure shows the individual contribution of each of the 40 discriminant metabolic features in the computed PCa metabolic score, i.e., the weight of each discriminant metabolite in the classification.

Table 2.3: IVDMIA performance for identified metabolites.

Feature Subpanel	Accuracy (%)	Specificity (%)	Sensitivity (%)	Discriminant Features (#)	Discriminant Feature Codes
Identified in Panel G by HRAM MS, and MS/MS	85.0	80.3	88.3	13	60, 36, 84, 71, 157, 176, 55, 343, 429, 384, 409, 386, 173
Identified in Panel A by HRAM MS, and MS/MS	91.1	91.3	90.9	31	60, 36, 84, 71, 157, 176, 55, 343, 429, 384, 409, 386, 173, 223, 43, 63, 376, 250, 211, 107, 214, 76, 444, 174, 128, 398, 93, 153, 364, 21, 242
Identified in Panel A by HRAM MS, and MS/MS and confirmed chromatographically with standards	76.3	70.6	79.9	10	60, 36, 71, 384, 43, 211, 76, 174, 128, 153
Identified in Panel A by HRAM MS, and MS/MS with xenobiotics and marker 63 excluded	90.2	90.7	89.7	28	60, 36, 84, 71, 157, 176, 55, 343, 429, 384, 409, 386, 173, 223, 43, 376, 250, 211, 107, 214, 76, 444, 174, 128, 398, 93, 153, 242

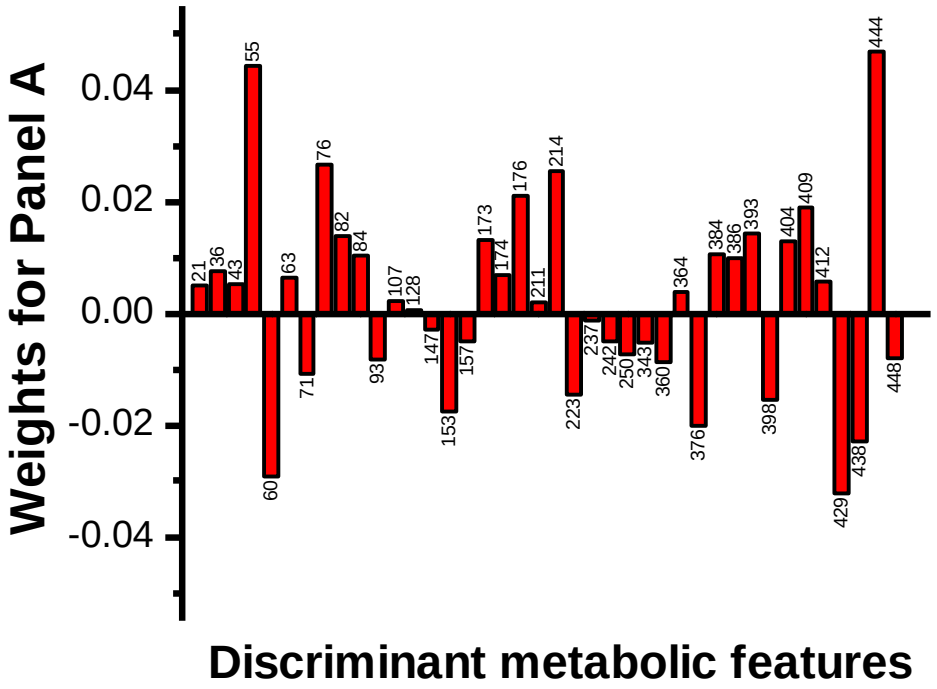

Figure 2.7: Weights for the 40 discriminant metabolic features in panel A. Metabolic features are labeled with their codes.

It is interesting to note that some features with high weights in the SVM model, such as feature 60, 444, 409, or 429, also have large absolute values in the PC3 loadings plot (Figure 2.5b). Figure 2.8 shows a comparison of the different sets of weights for the different panels described in Table 2.2, sorted from the largest to lowest value in panel A and expanded to panels B-G. The figure shows that the sign of the weights generally remained the same across the panels, in agreement with the fact that accuracy, specificity and sensitivity were highly conserved even after restricting the presence of discriminant features to those present in a majority of the patients within the cohort. It was seen that for the most restrictive panels, those features with weights equal to zero, i.e., those that do not contribute to the panels, are those with lower weights in panel A.

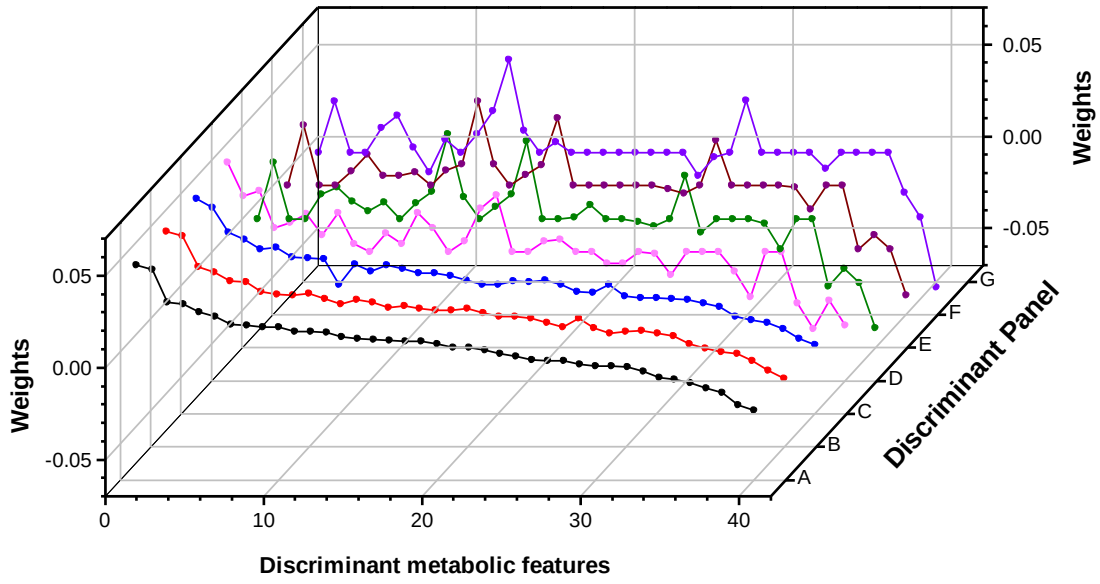

Figure 2.8: Weights for the discriminant metabolic features from panels A-G (indicated in Table 2.2) obtained by the classification model using the total cohort.

2.7 Identification of Metabolites Used in the *In Vitro* Diagnostic Multivariate Assay

Once the robustness of the model was established, chemical identification of the 40 discriminant metabolic features was attempted. Figure 2.9 exemplifies the procedure utilized for identification of feature 60. Figures 2.9a and 2.9b show the different base peak intensity (BPI) chromatograms obtained for serum samples of a typical PCa patient and healthy individual. As differences between metabolomes, and the corresponding features in the BPI chromatograms arise both from the presence of the disease and also from differences in diet, lifestyle, and the numerous other factors,[25] chemical identification of endogenous metabolites was attempted only for the 40 discriminant metabolic features.

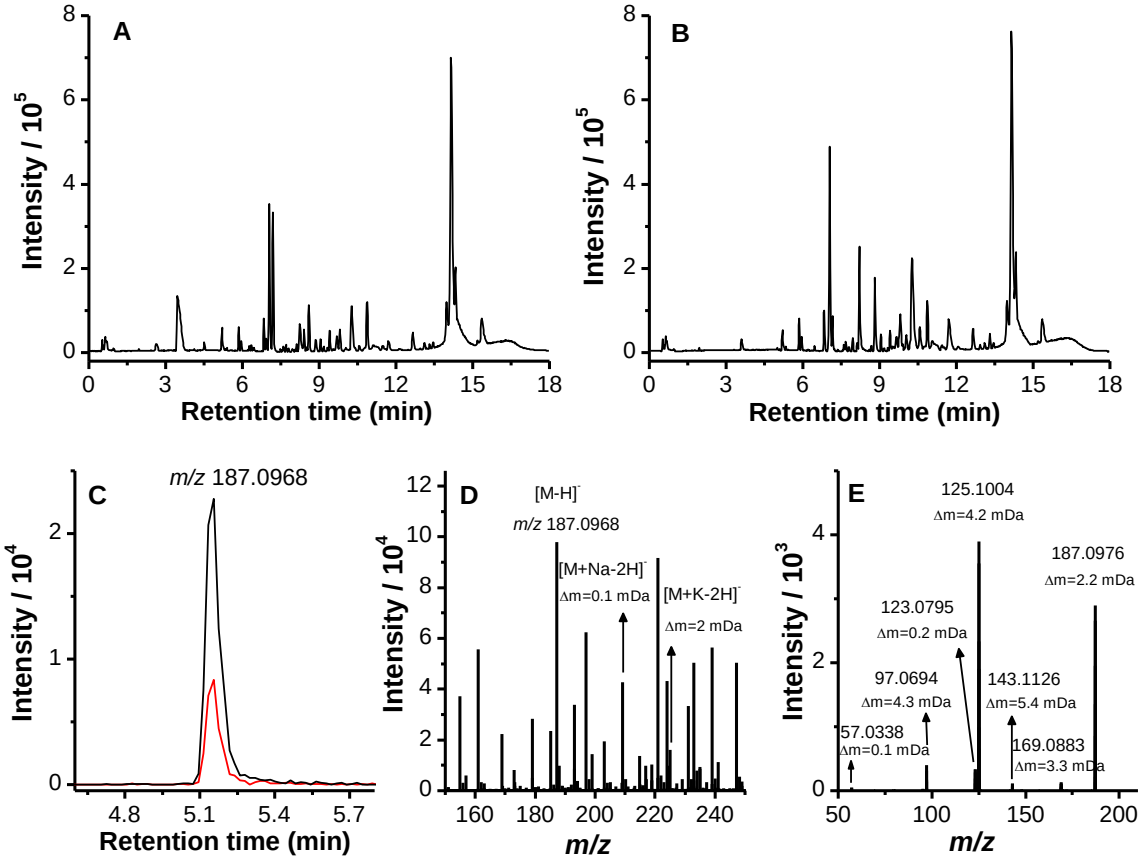

Figure 2.9: Base peak intensity chromatograms obtained for typical serum samples from a patient with PCa (A) and a healthy individual (B). (C): Extracted ion chromatogram for m/*z* 187.0968 ± 0.0050 generated from a PCa patient sample (red line) and a healthy individual (black line). These were generated from the data shown in A and B, respectively. (D): Adduct ion analysis for discriminant feature at m/*z* 187.0968. Mass errors are calculated with respect to the theoretical values for azelaic acid ($C_9H_{16}O_4$). Tandem MS spectrum for the m/*z* 187.0968 precursor ion using a collision cell voltage of 15 V. The matching of tandem MS fragmentation patterns between the experimental spectrum and the metabolite candidate is illustrated by the mass errors calculated as differences with the values in the METLIN database.

The high resolving power of the time-of-flight analyzer used allowed generating highly-selective extracted ion chromatograms for each discriminant feature, as illustrated in Figure 2.9c. Adduct ion analysis (Figure 2.9d) was used to ensure the unambiguous assignment of the signal of interest in the electrospray ionization mass spectrum, and the isotopic pattern and accurate masses were used to generate a list of possible candidate elemental formulae that were searched against databases. Moreover, UPLC-MS/MS experiments were performed to confirm the identities of these candidate metabolites responsible for classification. Tandem MS spectra were compared to those in databases or literature, and fragmentation patterns were manually analyzed as well (Figure 2.9e). Finally, standards of all commercially-available metabolites were subject to UPLC-MS and MS/MS to verify the identity of the candidates by retention time and mass spectral matching. Of the 40 spectral features found in panel A, 31 were identified by HRAM MS and MS/MS, with 10 further confirmed chromatographically by standards. The set of 31 metabolites provided 90.9% sensitivity, 91.3% specificity, and 91.1% accuracy; whereas the 10 differential metabolites confirmed by standards, when considered alone, provided 79.9% sensitivity, 70.6% specificity, and 76.3% accuracy (Table 2.3). It should be noted that among the 31 identified metabolites, 1-α-amino-1H-pyrrole-1-hexanoic acid (feature code 63) had the highest mass error (11.4 mDa), and its identity should be viewed as tentative. However, a classification model built using the set of 30 metabolites excluding feature 63 still provided 92.8% sensitivity, 89.2% specificity and 91.2% accuracy.

2.8 Biological Relevance of the *In Vitro* Diagnostic Multivariate Assay Metabolites

Table 2.4 summarizes the results from the chemical identification workflow described above for the 40 discriminant features. Those metabolites with chromatographic identity confirmation by retention time matching with standards are shown in bold, and can be therefore viewed as the ones with the higher confidence in the panel. Several discriminant metabolites were identified as fatty acids, amino acids, lysophospholipids, and bile acids, suggesting alterations in their respective metabolism. Previous findings have shown abnormality in fatty acid,[26] and amino acid[12, 27, 28] metabolism in PCa patients. Alterations in fatty acid metabolism through an enhanced β-oxidation pathway have been suggested to provide bioenergy for abnormal cell proliferation.[26] Among the different lysophospholipids identified that may play a role in cell signaling,[29] lysoPC(18:2) and lysoPC(18:0) have been reported as biomarkers for PCa detection within a panel of plasma lipids.[19] Uric acid has also been suggested to be a disease risk marker due to its pro-inflammatory properties,[30, 31] and a prospective epidemiological study demonstrated positive association between serum uric acid levels and risk of PCa development.[32] In addition, elevated concentrations of serum uric acid are often found due to tumor lysis syndrome observed as a result of cancer therapy.[33] Interestingly, indoxyl sulfate, a toxic product of dietary tryptophan metabolism that accumulates in the blood of patients with impaired renal function,[34] was also identified among the 40 discriminant features.

Table 2.4: Results for the chemical identification workflow for various discriminant features. Metabolites confirmed by retention time matching with commercially-available standards are highlighted in bold font.

Feature Code	Retention time (min)	m/z	Ion type	Elemental Formula	Theoretical m/z	Δm (mDa)	Tentative Metabolite Identification	Ref.	Panel
60	**5.10**	**187.0970**	**$[M-H]^-$**	**$C_9H_{16}O_4$**	**187.0970**	**0.0**	**nonanedioic acid (azelaic acid)**	**35**	**G**
36	**0.63**	**167.0206**	**$[M-H]^-$**	**$C_5H_4N_4O_3$**	**167.0205**	**0.1**	**uric acid**	**30-32**	**G**
71	**1.95**	**203.0817**	**$[M-H]^-$**	**$C_{11}H_{12}N_2O_2$**	**203.0821**	**0.4**	**tryptophan**	**27, 36**	**G**
384	**11.70**	**508.3403**	**$[M-CH_3]^-$**	**$C_{26}H_{54}NO_7P$**	**508.3403**	**0.0**	**lysoPC(18:0/0:0)**	**19, 37**	**G**
84	8.41	223.1331	$[M-H]^-$	$C_{13}H_{20}O_3$	223.1334	0.3	13-oxo-9,11-tridecadienoic acid	38	G
157	7.06	273.1703	$[M-H]^-$	$C_{14}H_{26}O_5$	273.1702	0.1	3-hydroxytetradecanedioic acid	39, 40	G
176	7.61	287.1854	$[M-H]^-$	$C_{15}H_{28}O_5$	287.1858	0.4	6-hydroxypentadecanedioic acid		G
55	5.21	185.0812	$[M-H]^-$	$C_9H_{14}O_4$	185.0814	0.2	5-(2-methylpropyl)-2-oxooxolane-3-carboxylic acid 5-butyl-2-oxooxolane-3-carboxylic acid	41	G
343	9.77	476.2772	$[M-H]^-$	$C_{23}H_{44}NO_7P$	476.2777	0.5	lysoPE(0:0/18:2) lysoPE(18:2/0:0)	42	G
429	9.80	578.3450	$[M+CH_3COO]^-$	$C_{26}H_{50}NO_7P$	578.3458	0.8	lysoPC(18:2/0:0)	19	G
409	5.46	541.2639	$[M-H]^-$	$C_{27}H_{42}O_{11}$	541.2649	1.0	cortolone-3-glucuronide	43, 44	G
386	6.92	511.2900	$[M-H]^-$	$C_{27}H_{44}O_9$	511.2907	0.7	pregnanetriol glucuronide		G
173	8.19	285.1920	$[M-H]^-$	$C_{19}H_{26}O_2$	285.1855	6.5	androstenedione	45	G

Table 2.4 (continued).

393	8.12	517.3015		-	-	-	-		G
438	7.04	600.2572		-	-	-	-		G
147	0.55	266.8028		-	-	-	-		G
82	8.12	215.1281		-	-	-	-		G
43	**9.56**	**171.1383**	**[M-H]$^-$**	**$C_{10}H_{20}O_2$**	**171.1385**	**0.2**	**decanoic acid (capric acid)**		**F**
223	6.77	331.1753	[M-H]$^-$	$C_{16}H_{28}O_7$	331.1757	0.4	menthol glucuronide citronellol glucuronide	46-49	F
63	7.19	195.1020	[M-H]$^-$	$C_{10}H_{16}N_2O_2$	195.1134	11.4	l-α-amino-1H-pyrrole-1-hexanoic acid		F
376	9.63	504.3081	[M-CH_3]$^-$	$C_{26}H_{50}NO_7P$	504.309	0.9	lysoPC(0:0/18:2)*	42	F
412	8.86	545.3323		-	-	-	-		F
211	**4.06**	**311.1387**	**[M-H]$^-$**	**$C_{18}H_{20}N_2O_3$**	**311.1396**	**0.9**	**phenylalanyl phenylalanine**	**28**	**E**
250	5.70	383.1521	[M-H]$^-$	$C_{19}H_{28}O_6S$	383.1528	0.7	3β,16α-dihydroxyandrostenone sulfate		E
107	5.40	245.0480	[M-H]$^-$	$C_{10}H_{14}O_5S$	245.0484	0.4	2-tert-butyl-1,4-benzenediol sulfate	50	E
76	**2.64**	**212.0016**	**[M-H]$^-$**	**$C_8H_7NO_4S$**	**212.0018**	**0.2**	**indoxyl sulfuric acid**	**12, 51**	**D**
214	9.87	311.2211	[M-H]$^-$	$C_{18}H_{32}O_4$	311.2222	1.1	9,10-dihydroxy-12Z,15Z-octadecadienoic acid (9,10-DiHODE) 12,13-dihydroxy-9Z,15Z-octadecadienoic acid (12,13-DiHODE) 15,16-dihydroxy-9Z,12Z-octadecadienoic acid (15,16-DiHODE)	52, 53	D

Table 2.4 (continued).

444	6.82	613.3583	$[M-H]^-$	$C_{32}H_{54}O_{11}$	613.3588	0.5	27-nor-5β-cholestane-3α,7α,12α,24,25-pentol glucuronide	54, 55	D
174	**9.35**	**285.2059**	**$[M-H]^-$**	**$C_{16}H_{30}O_4$**	**285.2066**	**0.7**	**hexadecanedioic acid**	**56**	**C**
128	**2.69**	**263.1023**	**$[M-H]^-$**	**$C_{13}H_{16}N_2O_4$**	**263.1032**	**0.9**	**phenylacetylglutamine**	**57, 58**	**C**
153	**14.80**	**269.2475**	**$[M-H]^-$**	**$C_{17}H_{34}O_2$**	**269.2481**	**0.6**	**heptadecanoic acid**	**37**	**C**
398	7.06	528.2630	$[M-H]^-$	$C_{26}H_{43}NO_8S$	528.2631	0.1	n-[(3α,5β,7β)-7-hydroxy-24-oxo-3-(sulfooxy)cholan-24-yl]-glycine n-[(3α,5β,7α)-3-hydroxy-24-oxo-7-(sulfooxy)cholan-24-yl]-glycine glycochenodeoxycholate-3-sulfate	59	C
93	6.36	229.0534	$[M-H]^-$	$C_{10}H_{14}O_4S$	229.0535	0.1	5-isopropyl-2-methylphenol sulfate (carvacrol sulfate)	60	C
360	8.16	489.2692		-	-	-	-		C
448	8.51	621.3273		-	-	-	-		C
364	5.57	495.2228	$[M-H]^-$	$C_{25}H_{36}O_{10}$	495.2230	0.2	5'-carboxy-α-chromanol glucuronide	61	B
21	5.16	144.0471	$[M-H]^-$	C_9H_7NO	144.0449	2.2	indole-3-carboxaldehyde	62, 63	B
404	7.28	537.2501		-	-	-	-		B
242	7.66	369.1740	$[M-H]^-$	$C_{19}H_{30}O_5S$	369.1736	0.4	androsterone sulfate 5α-dihydrotestosterone sulfate etiocholanolone sulfate	64-66	A

Table 2.4 (continued).

237	11.34	365.2680	-	-	-	-	A

Abbreviations: lysoPC: lysophosphatidylcholine; lysoPE: lysophosphatidylethanolamine
*not in HMDB

The reason behind elevated indoxyl sulfate in serum of PCa patients is not yet fully understood; nevertheless, this nephrotoxic metabolite likely contributes to the disease or its complications *via* multiple mechanisms, including enhanced oxidative stress due to decreased levels of glutathione.[51]

Perhaps the most salient finding resulting from the chemical identification workflow is that many differentiating metabolites belong to the steroid hormone biosynthesis pathway. The pathway supplies androgens[64-66] such as testosterone and 5α-dihydrotestosterone, to support the growth of androgen-dependent PCa.[67] An average increase of pregnanetriol and androstenedione concentrations in PCa serum suggests that there is a metabolic alteration of the steroid pathway that mimics congenital adrenal hyperplasia (CAH), a metabolic disease that is accompanied by androgen excess due to the diversion of 17-hydroxyprogesterone into the pathway for androgen biosynthesis.[68, 69] In addition, the average decrease of azelaic acid concentration in serum of PCa patients, an inhibitor of 5α-reductase,[70] suggests the disinhibition of 5α-reductase, an enzyme that catalyzes the synthesis of highly active androgen 5α-dihydrotestosterone to support PCa growth. Indeed, azelaic acid, which has a large contribution in the models, has been postulated to be a potential antitumoral agent.[35]

Table 2.4 also shows the identification of several xenobiotics that can be grouped into two classes according to their origin. Menthol, citronellol, carvacrol, and t-butylhydroquinone are most likely related to food components. Assuming that both PCa patients and healthy individuals were equally exposed, on average, to food components/additives, their different metabolism could explain the different levels of

these xenometabolites in serum. For example, the terpenoids menthol, carvacrol and citronellol are metabolized by CYP2A6,[71, 72] which is also involved in steroid metabolism. As a result, average lower concentrations of these terpenoids relative to healthy individuals may be suggestive of higher activity of CYP2A6 in PCa patients, supporting inclusion of these xenometabolites in the models. The second group of xenobiotics comprises indole-3-carboxaldehyde and 5'-carboxy-α-chromanol glucuronide, which could possibly result from the consumption of dietary supplements used by cancer patients. Self-medicating with an over-the-counter indole-3-carbinol (I3C) supplement may explain the increased average concentration of indole-3-carboxaldehyde in PCa serum.[62] Indeed, indole-3-carboxaldehyde demonstrated activity against prostate cancer in both *in vitro* and *in vivo* models.[63] Similarly, α-tocopherol, a form of vitamin E and a precursor of 5'-carboxy-α-chromanol glucuronide, have been suggested to influence the development of PCa due to their antioxidant activity.[61] As humans do not normally produce indole-3-carbaldehyde or 5'-carboxy-α-chromanol, and their consideration in the models may reflect dietary supplementation differences rather than endogenous metabolic differences, PCa detection was attempted using 28 of the 31 identified metabolites, excluding from the SVM classification model two metabolites which might result from dietary supplementation and one metabolite with highest mass error (1-α-amino-1H-pyrrole-1-hexanoic acid). This modified classification model provided 89.7% sensitivity, 90.7% specificity, and 90.2% accuracy (Table 2.3), indicating the three excluded metabolites had little effect on the overall assay performance, as supported by their low weights in panel A (Figure 2.7 and Figure 2.8).

2.9 Conclusion

The study presented here shows the combined application of UPLC-MS/MS and machine learning methods to develop a metabolite-based IVDMIA that predicts the presence of PCa in serum samples with high classification sensitivity, specificity and accuracy. A panel of 40 metabolic spectral features was found to be differential with 92.1% sensitivity, 94.3% specificity, and 93.0% accuracy. Of further significance, the detection performance of the IVDMIA was proven to be higher than the prevalent PSA test; highlighting that a combination of multiple discriminant features yields higher predictive power for PCa detection than the univariate analysis of a single marker. Within the discriminant panel, 31 metabolites were identified by HRAM MS and MS/MS, with 10 further confirmed chromatographically by standards. Fatty acids, amino acids, lysophospholipids, and bile acids have been identified among the discriminant metabolites, suggesting alterations in their metabolism. Additionally, several metabolites were mapped to the steroid hormone biosynthesis pathway. These observations demonstrate some of the plausible metabolic alterations in PCa, and provide further insight into the biological pathway changes associated with the disease. The combination of multiple metabolites that yield a single, patient-specific result for disease detection is the strength of the IVDMIA presented here. When the assay is based on the 28 identified disease-related metabolites, PCa can still be detected with 89.7% sensitivity, 90.7% specificity, and 90.2% accuracy. If higher throughput analysis, and lower analysis cost and complexity are desired, 13 metabolites that were found to be present in 90% of the entire sample cohort would still provide high classification sensitivity (88.3%), specificity

(80.3%), and accuracy (85.0%) for cancerous and healthy samples. Therefore, this assay shows promise towards its implementation in the clinical laboratory setting once it is fully validated by the examination of a larger patient cohort through targeted assays.

2.10 References

(1) Society, A. C., *Cancer Facts & Figures 2013*. American Cancer Society: Atlanta, 2013.

(2) Society, A. C., *Cancer Facts & Figures 2010*. American Cancer Society: Atlanta, 2010.

(3) Nadler, R. B.; Humphrey, P. A.; Smith, D. S.; Catalona, W. J.; Ratliff, T. L., Effect of inflammation and benign prostatic hyperplasia on elevated serum prostate-specific antigen levels. *J. Urol.* **1995,** *154* (2), 407-413.

(4) Tombal, B., Over- and underdiagnosis of prostate cancer: the dangers. *Eur. Urol. Suppl.* **2006,** *5* (6), 511-513.

(5) Bickers, B.; Aukim-Hastie, C., New molecular biomarkers for the prognosis and management of prostate cancer - the post PSA era. *Anticancer Res.* **2009,** *29* (8), 3289-3298.

(6) Heijnsdijk, E. A. M.; der Kinderen, A.; Wever, E. M.; Draisma, G.; Roobol, M. J.; de Koning, H. J., Overdetection, overtreatment and costs in prostate-specific antigen screening for prostate cancer. *Br. J. Cancer* **2009,** *101* (11), 1833-1838.

(7) Draisma, G.; Etzioni, R.; Tsodikov, A.; Mariotto, A.; Wever, E.; Gulati, R.; Feuer, E.; de Koning, H., Lead time and overdiagnosis in Prostate-Specific Antigen screening: importance of methods and context. *J. Natl. Cancer Inst.* **2009,** *101* (6), 374-383.

(8) Thompson, I. M.; Pauler, D. K.; Goodman, P. J.; Tangen, C. M.; Lucia, M. S.; Parnes, H. L.; Minasian, L. M.; Ford, L. G.; Lippman, S. M.; Crawford, E. D.; Crowley, J. J.; Coltman, C. A., Prevalence of prostate cancer among men with a prostate-specific antigen level ≤4.0 ng per milliliter. *New Engl. J. Med.* **2004,** *350* (22), 2239-2246.

(9) Trock, B. J., Application of metabolomics to prostate cancer. *Urol. Oncol.-Semin. Orig. Investig.* **2011,** *29* (5), 572-581.

(10) Chace, D. H.; Kalas, T. A.; Naylor, E. W., The application of tandem mass spectrometry to neonatal screening for inherited disorders of intermediary metabolism. *Annu. Rev. Genom. Hum. Genet.* **2002,** *3*, 17-45.

(11) Piraud, M.; Vianey-Saban, C.; Petritis, K.; Elfakir, C.; Steghens, J. P.; Morla, A.; Bouchu, D., ESI-MS/MS analysis of underivatised amino acids: a new tool for the

diagnosis of inherited disorders of amino acid metabolism. Fragmentation study of 79 molecules of biological interest in positive and negative ionisation mode. *Rapid Commun. Mass Spectrom.* **2003,** *17* (12), 1297-1311.

(12) Sreekumar, A.; Poisson, L. M.; Rajendiran, T. M.; Khan, A. P.; Cao, Q.; Yu, J. D.; Laxman, B.; Mehra, R.; Lonigro, R. J.; Li, Y.; Nyati, M. K.; Ahsan, A.; Kalyana-Sundaram, S.; Han, B.; Cao, X. H.; Byun, J.; Omenn, G. S.; Ghosh, D.; Pennathur, S.; Alexander, D. C.; Berger, A.; Shuster, J. R.; Wei, J. T.; Varambally, S.; Beecher, C.; Chinnaiyan, A. M., Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. *Nature* **2009,** *457* (7231), 910-914.

(13) Jentzmik, F.; Stephan, C.; Miller, K.; Schrader, M.; Erbersdobler, A.; Kristiansen, G.; Lein, M.; Jung, K., Sarcosine in urine after digital rectal examination fails as a marker in prostate cancer detection and identification of aggressive tumours. *Eur. Urol.* **2010,** *58* (1), 12-18.

(14) Schalken, J. A., Is urinary sarcosine useful to identify patients with significant prostate cancer? The trials and tribulations of biomarker development. *Eur. Urol.* **2010,** *58* (1), 19-20.

(15) Struys, E. A.; Heijboer, A. C.; van Moorselaar, J.; Jakobs, C.; Blankenstein, M. A., Serum sarcosine is not a marker for prostate cancer. *Ann. Clin. Biochem.* **2010,** *47* (Pt 3), 282.

(16) Swanson, M. G.; Vigneron, D. B.; Tabatabai, Z. L.; Males, R. G.; Schmitt, L.; Carroll, P. R.; James, J. K.; Hurd, R. E.; Kurhanewicz, J., Proton HR-MAS spectroscopy and quantitative pathologic analysis of MRI/3D-MRSI-targeted postsurgical prostate tissues. *Magn. Reson. Med.* **2003,** *50* (5), 944-954.

(17) Swanson, M. G.; Zektzer, A. S.; Tabatabai, Z. L.; Simko, J.; Jarso, S.; Keshari, K. R.; Schmitt, L.; Carroll, P. R.; Shinohara, K.; Vigneron, D. B.; Kurhanewicz, J., Quantitative analysis of prostate metabolites using H-1 HR-MAS spectroscopy. *Magn. Reson. Med.* **2006,** *55* (6), 1257-1264.

(18) Thysell, E.; Surowiec, I.; Hornberg, E.; Crnalic, S.; Widmark, A.; Johansson, A. I.; Stattin, P.; Bergh, A.; Moritz, T.; Antti, H.; Wikstro, P., Metabolomic characterization of human prostate cancer bone metastases reveals increased levels of cholesterol. *PLoS One* **2010,** *5* (12), e14175.

(19) Zhou, X. C.; Mao, J. H.; Ai, J. M.; Deng, Y. P.; Roth, M. R.; Pound, C.; Henegar, J.; Welti, R.; Bigler, S. A., Identification of plasma lipid biomarkers for prostate cancer by lipidomics and bioinformatics. *PLoS One* **2012,** *7* (11), e48889.

(20) Brereton, R. G.; Lloyd, G. R., Support Vector Machines for classification and regression. *Analyst* **2010,** *135* (2), 230-267.

(21) Guan, W.; Zhou, M.; Hampton, C.; Benigno, B.; Walker, L. D.; Gray, A.; McDonald, J.; Fernandez, F., Ovarian cancer detection from metabolomic liquid chromatography/mass spectrometry data by support vector machines. *BMC Bioinformatics* **2009,** *10* (1), 259.

(22) Smith, C. A.; O'Maille, G.; Want, E. J.; Qin, C.; Trauger, S. A.; Brandon, T. R.; Custodio, D. E.; Abagyan, R.; Siuzdak, G., METLIN - a metabolite mass spectral database. *Ther. Drug Monit.* **2005,** *27* (6), 747-751.

(23) Wishart, D. S.; Jewison, T.; Guo, A. C.; Wilson, M.; Knox, C.; Liu, Y. F.; Djoumbou, Y.; Mandal, R.; Aziat, F.; Dong, E.; Bouatra, S.; Sinelnikov, I.; Arndt, D.; Xia, J. G.; Liu, P.; Yallou, F.; Bjorndahl, T.; Perez-Pineiro, R.; Eisner, R.; Allen, F.; Neveu, V.; Greiner, R.; Scalbert, A., HMDB 3.0-the human metabolome database in 2013. *Nucleic Acids Res.* **2013,** *41* (D1), D801-D807.

(24) Horai, H.; Arita, M.; Kanaya, S.; Nihei, Y.; Ikeda, T.; Suwa, K.; Ojima, Y.; Tanaka, K.; Tanaka, S.; Aoshima, K.; Oda, Y.; Kakazu, Y.; Kusano, M.; Tohge, T.; Matsuda, F.; Sawada, Y.; Hirai, M. Y.; Nakanishi, H.; Ikeda, K.; Akimoto, N.; Maoka, T.; Takahashi, H.; Ara, T.; Sakurai, N.; Suzuki, H.; Shibata, D.; Neumann, S.; Iida, T.; Tanaka, K.; Funatsu, K.; Matsuura, F.; Soga, T.; Taguchi, R.; Saito, K.; Nishioka, T., MassBank: a public repository for sharing mass spectral data for life sciences. *J. Mass Spectrom.* **2010,** *45* (7), 703-714.

(25) Nicholson, J. K.; Lindon, J. C., Systems biology - metabonomics. *Nature* **2008,** *455* (7216), 1054-1056.

(26) Liu, Y., Fatty acid oxidation is a dominant bioenergetic pathway in prostate cancer. *Prostate Cancer Prostatic Dis.* **2006,** *9* (3), 230-234.

(27) Ebenezar, J.; Pu, Y.; Wang, W. B.; Liu, C. H.; Alfano, R. R., Stokes shift spectroscopy pilot study for cancerous and normal prostate tissues. *Appl. Opt.* **2012,** *51* (16), 3642-3649.

(28) Fu, Y. M.; Lin, H.; Liu, X.; Fang, W.; Meadows, G. G., Cell death of prostate cancer cells by specific amino acid restriction depends on alterations of glucose metabolism. *J. Cell. Physiol.* **2010,** *224* (2), 491-500.

(29) Peyruchaud, O., Novel implications for lysophospholipids, lysophosphatidic acid and sphingosine 1-phosphate, as drug targets in cancer. *Anticancer Agents Med. Chem.* **2009,** *9* (4), 381-391.

(30) Brys, M.; Morel, A.; Forma, E.; Krzeslak, A.; Wilkosz, J.; Rozanski, W.; Olas, B., Relationship of urinary isoprostanes to prostate cancer occurence. *Mol. Cell. Biochem.* **2013,** *372* (1-2), 149-153.

(31) Fini, M.; Elias, A.; Johnson, R.; Wright, R., Contribution of uric acid to cancer risk, recurrence, and mortality. *Clin. Transl Med.* **2012,** *1* (1), 16.

(32) Kolonel, L.; Yoshizawa, C.; Nomura, A.; Stemmermann, G., Relationship of serum uric acid to cancer occurrence in a prospective male cohort. *Cancer Epidemiol. Biomarkers Prev.* **1994,** *3* (3), 225 - 228.

(33) Tsimberidou, A. M.; Keating, M. J., Hyperuricemic Syndromes in Cancer Patients. In *Hyperuricemic Syndromes: Pathophysiology and Therapy*, Ronco, C.; Rodeghiero, F., Eds. Karger Publishers: Basel, 2005; Vol. 147, pp 47-60.

(34) Barreto, F. C.; Barreto, D. V.; Liabeuf, S.; Meert, N.; Glorieux, G.; Temmar, M.; Choukroun, G.; Vanholder, R.; Massy, Z. A.; European Uremic Toxin Work Grp, E. U. T., Serum indoxyl sulfate is associated with vascular disease and mortality in chronic kidney disease patients. *Clin. J. Am. Soc. Nephrol.* **2009,** *4* (10), 1551-1558.

(35) Breathnach, A. S., Azelaic acid: potential as a general antitumoural agent. *Med. Hypotheses* **1999,** *52* (3), 221-226.

(36) Prendergast, G. C., Cancer: why tumours eat tryptophan. *Nature* **2011,** *478* (7368), 192-194.

(37) Crowe, F. L.; Allen, N. E.; Appleby, P. N.; Overvad, K.; Aardestrup, I. V.; Johnsen, N. F.; Tjonneland, A.; Linseisen, J.; Kaaks, R.; Boeing, H.; Kroger, J.; Trichopoulou, A.; Zavitsanou, A.; Trichopoulos, D.; Sacerdote, C.; Palli, D.; Tumino, R.; Agnoli, C.; Kiemeney, L. A.; Bueno-de-Mesquita, H. B.; Chirlaque, M. D.; Ardanaz, E.; Larranaga, N.; Quiros, J. R.; Sanchez, M. J.; Gonzalez, C. A.; Stattin, P.; Hallmans, G.; Bingham, S.; Khaw, K. T.; Rinaldi, S.; Slimani, N.; Jenab, M.; Riboli, E.; Key, T. J., Fatty acid composition of plasma phospholipids and risk of prostate cancer in a case-control analysis nested within the European Prospective Investigation into Cancer and Nutrition. *Am. J. Clin. Nutr.* **2008,** *88* (5), 1353-63.

(38) Glasgow, W.; Eling, T., Structural Requirements for Enhancement of EGF-Dependent DNA Synthesis by Oxygenated Metabolltes of Linoleic Acid. In *Eicosanoids and Other Bioactive Lipids in Cancer, Inflammation, and Radiation Injury 2*, Honn, K.; Nigam, S.; Marnett, L., Eds. Springer US: New York City, 1997; Vol. 400, pp 507-512.

(39) Tserng, K. Y.; Jin, S. J., Metabolic origin of urinary 3-hydroxy dicarboxylic acids. *Biochemistry* **1991,** *30* (9), 2508-14.

(40) Zha, S.; Ferdinandusse, S.; Hicks, J. L.; Denis, S.; Dunn, T. A.; Wanders, R. J.; Luo, J.; De Marzo, A. M.; Isaacs, W. B., Peroxisomal branched chain fatty acid β-oxidation pathway is upregulated in prostate cancer. *Prostate* **2005,** *63* (4), 316-323.

(41) Pettersson, A.; Kasperzyk, J. L.; Kenfield, S. A.; Richman, E. L.; Chan, J. M.; Willett, W. C.; Stampfer, M. J.; Mucci, L. A.; Giovannucci, E. L., Milk and Dairy Consumption among Men with Prostate Cancer and Risk of Metastases and Prostate Cancer Death. *Cancer Epidemiol., Biomarkers Prev.* **2012,** *21* (3), 428-436.

(42) Sun, H.; Zhang, A. H.; Yan, G. L.; Piao, C. Y.; Li, W. Y.; Sun, C.; Wu, X. H.; Li, X. H.; Chen, Y.; Wang, X. J., Metabolomic analysis of key regulatory metabolites in Hepatitis C virus-infected tree shrews. *Mol. Cell. Proteomics* **2013,** *12* (3), 710-719.

(43) Cho, H. J.; Kim, J. D.; Lee, W. Y.; Chung, B. C.; Choi, M. H., Quantitative metabolic profiling of 21 endogenous corticosteroids in urine by liquid chromatography-triple quadrupole-mass spectrometry. *Anal. Chim. Acta* **2009,** *632* (1), 101-8.

(44) Schatzl, G.; Reiter, W. J.; Thurridl, T.; Waldmuller, J.; Roden, M.; Soregi, S.; Madersbacher, S., Endocrine patterns in patients with benign and malignant prostatic diseases. *Prostate* **2000,** *44* (3), 219-224.

(45) Barrett-Connor, E.; Garland, C.; McPhillips, J. B.; Khaw, K. T.; Wingard, D. L., A prospective, population-based study of androstenedione, estrogens, and prostatic cancer. *Cancer Res.* **1990,** *50* (1), 169-73.

(46) Li, Q.; Wang, X.; Yang, Z.; Wang, B.; Li, S., Menthol induces cell death via the TRPM8 channel in the human bladder cancer cell line T24. *Oncology* **2009,** *77* (6), 335-341.

(47) Lin, J. P.; Lu, H. F.; Lee, J. H.; Lin, J. G.; Hsia, T. C.; Wu, L. T.; Chung, J. G., (-)-Menthol inhibits DNA topoisomerases I, II alpha and beta and promotes NF-kappaB expression in human gastric cancer SNU-5 cells. *Anticancer Res.* **2005,** *25* (3B), 2069-74.

(48) Sacchetti, G.; Maietti, S.; Muzzoli, M.; Scaglianti, M.; Manfredini, S.; Radice, M.; Bruni, R., Comparative evaluation of 11 essential oils of different origin as functional antioxidants, antiradicals and antimicrobials in foods. *Food Chem.* **2005,** *91* (4), 621-632.

(49) Brito, R. G.; Guimaraes, A. G.; Quintans, J. S. S.; Santos, M. R. V.; De Sousa, D. P.; Badaue-Passos, D., Jr.; de Lucca, W., Jr.; Brito, F. A.; Barreto, E. O.; Oliveira, A. P.; Quintans, L. J., Jr., Citronellol, a monoterpene alcohol, reduces nociceptive and inflammatory activities in rodents. *J. Nat. Med.* **2012,** *66* (4), 637-644.

(50) Lau, A.; Villeneuve, N. F.; Sun, Z.; Wong, P. K.; Zhang, D. D., Dual roles of Nrf2 in cancer. *Pharmacol. Res.* **2008,** *58* (5-6), 262-270.

(51) Dou, L.; Jourde-Chiche, N.; Faure, V.; Cerini, C.; Berland, Y.; Dignat-George, F.; Brunet, P., The uremic solute indoxyl sulfate induces oxidative stress in endothelial cells. *J. Thromb. Haemost.* **2007,** *5* (6), 1302-8.

(52) Collett, G. P.; Betts, A. M.; Johnson, M. I.; Pulimood, A. B.; Cook, S.; Neal, D. E.; Robson, C. N., Peroxisome proliferator-activated receptor alpha is an androgen-responsive gene in human prostate and is highly expressed in prostatic adenocarcinoma. *Clin. Cancer. Res.* **2000,** *6* (8), 3241-3248.

(53) Shearer, G. C.; Harris, W. S.; Pedersen, T. L.; Newman, J. W., Detection of omega-3 oxylipins in human plasma and response to treatment with omega-3 acid ethyl esters. *J. Lipid Res.* **2010,** *51* (8), 2074-2081.

(54) Karlaganis, G.; Bremmelgaard, A.; Karlaganis, V.; Sjovall, J., Precursor of 27-nor-5beta-cholestane-3alpha,7alpha,12alpha,24,25-pentol in man *J. Steroid Biochem. Mol. Biol.* **1983,** *18* (6), 725-729.

(55) Krycer, J. R.; Brown, A. J., Cholesterol accumulation in prostate cancer: A classic observation from a modern perspective. *Biochim. Biophys. Acta-Rev. Cancer* **2013,** *1835* (2), 219-229.

(56) Li, F.; Qin, X.; Chen, H.; Qiu, L.; Guo, Y.; Liu, H.; Chen, G.; Song, G.; Wang, X.; Li, F.; Guo, S.; Wang, B.; Li, Z., Lipid profiling for early diagnosis and progression of colorectal cancer using direct-infusion electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. *Rapid Commun. Mass Spectrom.* **2013,** *27* (1), 24-34.

(57) Onishi, T.; Yamakawa, K.; Franco, O. E.; Suzuki, R.; Kawamura, J., p27(Kip1) is the key mediator of phenylacetate induced cell cycle arrest in human prostate cancer cells. *Anticancer Res.* **2000,** *20* (5A), 3075-3081.

(58) Shibahara, T.; Onishi, T.; Franco, O. E.; Arima, K.; Sugimura, Y., Down-regulation of Skp2 is correlated with p27-associated cell cycle arrest induced by phenylacetate in human prostate cancer cells. *Anticancer Res.* **2005,** *25* (3B), 1881-1888.

(59) Goldberg, A. A.; Titorenko, V. I.; Beach, A.; Sanderson, J. T., Bile acids induce apoptosis selectively in androgen-dependent and -independent prostate cancer cells. *PeerJ* **2013,** *1*, e122.

(60) Patel, B.; Shah, V. R.; Bavadekar, S. A., Anti-proliferative effects of carvacrol on human prostate cancer cell line, LNCaP. *FASEB J.* **2012,** *26*.

(61) Heinonen, O. P.; Albanes, D.; Virtamo, J.; Taylor, P. R.; Huttunen, J. K.; Hartman, A. M.; Haapakoski, J.; Malila, N.; Rautalahti, M.; Ripatti, S.; Maenpaa, H.; Teerenhovi, L.; Koss, L.; Virolainen, M.; Edwards, B. K., Prostate cancer and supplementation with alpha-tocopherol and beta-carotene: incidence and mortality in a controlled trial. *J. Natl. Cancer Inst.* **1998,** *90* (6), 440-446.

(62) Anderton, M. J.; Manson, M. M.; Verschoyle, R. D.; Gescher, A.; Lamb, J. H.; Farmer, P. B.; Steward, W. P.; Williams, M. L., Pharmacokinetics and tissue disposition of indole-3-carbinol and its acid condensation products after oral administration to mice. *Clin. Cancer. Res.* **2004,** *10* (15), 5233-5241.

(63) Souli, E.; Machluf, M.; Morgenstern, A.; Sabo, E.; Yannai, S., Indole-3-carbinol (I3C) exhibits inhibitory and preventive effects on prostate tumors in mice. *Food Chem. Toxicol.* **2008,** *46* (3), 863-870.

(64) Titus, M. A.; Schell, M. J.; Lih, F. B.; Tomer, K. B.; Mohler, J. L., Testosterone and dihydrotestosterone tissue levels in recurrent prostate cancer. *Clin. Cancer. Res.* **2005,** *11* (13), 4653-4657.

(65) Kliman, B.; Prout, G. R.; Maclaughlin, R. A.; Daly, J. J.; Griffin, P. P., Altered androgen metabolism in metastatic prostate cancer *J. Urol.* **1978,** *119* (5), 623-626.

(66) Zumoff, B.; Levin, J.; Strain, G. W.; Rosenfeld, R. S.; Oconnor, J.; Freed, S. Z.; Kream, J.; Whitmore, W. S.; Fukushima, D. K.; Hellman, L., Abnormal levels of plasma hormones in men with prostate-cancer - evidence toward a 2-disease theory. *Prostate* **1982,** *3* (6), 579-588.

(67) Knudsen, K. E.; Scher, H. I., Starving the addiction: new opportunities for durable suppression of AR signaling in prostate cancer. *Clin. Cancer. Res.* **2009,** *15* (15), 4792-4798.

(68) Horton, R.; Frasier, S. D., Androstenedione and its conversion to plasma testosterone in congenital adrenal hyperplasia *J. Clin. Invest.* **1967,** *46* (6), 1003-1009.

(69) Speiser, P. W.; White, P. C., Congenital adrenal hyperplasia. *New Engl. J. Med.* **2003,** *349* (8), 776-788.

(70) Passi, S.; Picardo, M.; De Luca, C.; Nazzaro-Porro, M., Mechanism of azelaic acid action in acne. *G. Ital. Dermatol. Venereol.* **1989,** *124* (10), 455-463.

(71) Dong, R. H.; Fang, Z. Z.; Zhu, L. L.; Ge, G. B.; Cao, Y. F.; Li, X. B.; Hu, C. M.; Yang, L.; Liu, Z. Y., Identification of CYP isoforms involved in the metabolism of thymol and carvacrol in human liver microsomes (HLMs). *Pharmazie* **2012,** *67* (12), 1002-1006.

(72) Miyazawa, M.; Marumoto, S.; Takahashi, T.; Nakahashi, H.; Haigou, R.; Nakanishi, K., Metabolism of (+)- and (-)-menthols by CYP2A6 in human liver microsomes. *J. Oleo Sci.* **2011,** *60* (3), 127-132.

CHAPTER 3. ULTRA PERFORMANCE LIQUID CHROMATOGRAPHY-MASS SPECTROMETRY SERUM METABOLOMICS DETECTION OF EARLY-STAGE OVARIAN CANCER

Adapted with permission from
Jones, C. M.; Monge, M. E.; Kim, J.; Matzuk, M. M.; Fernández, F. M., Metabolomic Serum Profiling Detects Early-Stage High-Grade Serous Ovarian Cancer in a Mouse Model. *J. Proteome Res.* **2015**, *14* (2), 917-927. Copyright © 2015 American Chemical Society.

Jones, C. M.; Gaul, D.; Long, T. Q.; Monge, M. E.; Walker, L. D.; McDonald, J. F.; Fernández, F. M., Ultra Performance Liquid Chromatography-Mass Spectrometry Characterization of Serum Metabolic Phenotypes of an Early-Stage Ovarian Cancer Pilot Patient Cohort. In Preparation.

This chapter describes research conducted by multiple persons. J. Kim collected blood serum samples from DKO and control mice. C. M. Jones and M. E. Monge optimized all sample preparation protocols and UPLC-MS analysis methods and processed the UPLC-MS data for the mouse model study. D. Gaul processed the UPLC-MS data for the human cohort study. T. Q. Long performed the support vector machine multivariate analysis for the human cohort study. C. M. Jones performed all PCA and PLS-DA multivariate analyses, conducted all UPLC-MS/MS and chemical standard validation experiments, and determined metabolite identities and biological functions.

3.1 Abstract

Ovarian cancer (OC) is the 5th leading cause of cancer-related deaths for U.S. women, yet it has the highest mortality rate amongst gynecological cancers. Non-specific symptoms, combined with a lack of early detection methods and highly specific biomarkers, contribute to late diagnosis and low 5-year survival rates; thus, an effective screening strategy for early diagnosis would be particularly advantageous since 5-year survival rates can be as high as 90%. Two approaches were taken to investigate metabolic patterns for early detection of this deadly disease. First, *Dicer-Pten* double knockout

(DKO) mice that phenocopy many of the features of metastatic high-grade serous carcinoma (HGSC) observed in women were studied. HGSC is the most common and deadliest subtype that results in 90% of OC deaths. Using ultra performance liquid chromatography-mass spectrometry (UPLC-MS), serum samples from 14 early-stage tumor (ET) DKO mice and 11 controls were analyzed in depth to screen for metabolic signatures capable of differentiating early-stage HGSC from controls. Iterative multivariate classification selected 18 metabolites that, when considered as a panel, yielded 100% accuracy, sensitivity, and specificity for classification. Altered metabolic pathways reflected in that panel included those of fatty acids, bile acids, glycerophospholipids, peptides, and some dietary phytochemicals. These alterations revealed impacts to cellular energy storage and membrane stability, as well as changes in defenses against oxidative stress, shedding new light on the metabolic alterations associated with early OC stages.

In the second approach, serum metabolic phenotypes of an early-stage OC pilot patient cohort were characterized. Serum samples were collected from 24 early-stage OC patients and 40 healthy women, and subsequently analyzed using UPLC-MS. Multivariate statistical analysis employing support vector machine (SVM) learning methods and recursive feature elimination (RFE) selected a panel of metabolites that differentiated between age-matched samples with 100% cross-validated accuracy, sensitivity, and specificity. This small pilot study demonstrated that metabolic phenotypes may be useful for detecting early-stage OC and, thus, supports conducting larger, more comprehensive studies.

3.2 Ovarian Cancer

3.2.1 Overview

Ovarian cancer (OC) is the 5th leading cause of cancer-related deaths for U.S. women.[1] In particular, high-grade serous carcinoma (HGSC), the subtype with the highest occurrence and mortality, is responsible for 90% of all ovarian cancer deaths, yet its origin and early progression are poorly understood.[2-4] Due to the unavailability of reliable screening tests in clinical practice and the asymptomatic course through early stages of the disease, the majority of ovarian cancer cases (68%), including most HGSCs (>95%), are diagnosed as advanced, metastatic disease with poor survival.[5, 6] The 5-year OC survival rate for all cases diagnosed during 2003-2009 was 44%.[7] When the cancer is confined to the ovary at diagnosis, however, the 5-year survival is over 90%.[2] Early detection is thus crucial in reducing ovarian cancer mortality.

3.2.2 Current Diagnostic Methodology

The conventional evaluation of OC patients includes physical examination, transvaginal ultrasonography, and measurement of levels of the serum tumor biomarker CA125. However, this marker is of limited utility since it can also be elevated by conditions unrelated to ovarian cancer, especially in premenopausal women.[8] Recent data[9, 10] have suggested that the OVA1[11] test, the first protein-based *In Vitro* Diagnostic Multivariate Index Assay (IVDMIA) approved by the FDA, may improve, along with physician clinical assessment, detection rates of malignancies among women with pelvic masses planning to undergo surgery. Still, OVA1 is not yet an OC screening or definitive

diagnostic test. Furthermore, whether this assay can detect OC at an early time point still remains unclear,[12] and a more effective screening strategy for early diagnosis would be particularly advantageous for patients.

3.2.3 *Dicer-Pten* Double-Knockout Mouse Model

Traditionally, OC has been thought to originate in the ovary. The fallopian tube, however, has recently been proposed as an alternate site of origin, especially in women carrying hereditary *BRCA* mutations.[13-15] A mouse model of HGSC where disease originates through this alternative route was therefore developed by conditionally disabling two critical genes, *Dicer* and *Pten* (*Dicer* $^{flox/flox}$ *Pten* $^{flox/flox}$ *Amhr2* $^{cre/+}$) in the fallopian tubes.[16] In these *Dicer-Pten* double-knockout (DKO) mice, HGSCs originate and progressively develop in the fallopian tube before spreading to the ovary, and then metastasize throughout the abdominal cavity, causing ascites, and eventually killing the mice. Besides replicating the clinical biology of human HGSC in that tumors are characterized by complex papillae and irregular glands forming slit-like spaces in addition to solid sheets of tumor cells with pleomorphic nuclei, prominent nucleoli, and elevated mitotic activity, disease in these DKO mice also shows close molecular similarities with human HGSCs, such as upregulated folate receptor 1 (*Folr1*), CA125 (*Muc16*), secreted phosphoprotein 1 (*Spp1*), and chemokine genes, therefore providing a simpler, better-controlled, model to study early-stage OC which could potentially be later translated to humans.

3.2.4 Metabolic Ovarian Cancer Detection

During the last decade, metabolomics has emerged as a promising discipline providing tools to investigate characteristic metabolic patterns of disease, with one of its goals being the discovery of biomarker panels for early diagnosis. Mass spectrometry (MS) and ^{1}H nuclear magnetic resonance (NMR) spectroscopy in combination with multivariate statistical analysis have been utilized to investigate ovarian cancer-induced metabolome alterations in urine,[17-20] plasma[21], serum[22-26], and tissues.[27-29] Li and collaborators, for example, identified L-tryptophan, lysoPC(18:3), lysoPC(14:0), and 2-piperidinone as plasma metabolites discriminating between epithelial ovarian cancer (EOC) patients and women with benign ovarian tumors.[21] Disruption to nucleotide, histidine, tryptophan, and mucin metabolism pathways[17, 21] and changes in amino acids involved in *de novo* purine nucleotide synthesis, have also been reported.[26] However, and despite these advances, no widely-accepted strategy for metabolome-based OC screening has yet emerged.[12, 30] Moreover, only a very small number of studies have focused on early-stage OC detection. Odunsi and collaborators[24], for example, used NMR spectroscopy to investigate metabolome changes in early-stage patients; independent validation of their reported predictive statistical model resulted in 95% specificity, 68% sensitivity, and an area under the Receiver Operator Characteristic Curve (AUC) of 0.949. Additionally, Xu and collaborators suggested the serum metabolite 27-nor-5β-cholestane-3,7,12,24,25 pentol glucuronide as a potential biomarker for stage I OC (specificity: 77%; sensitivity: 70%; AUC: 0.750).[25] A hindrance plaguing this field has been the poor understanding of early tumor development mechanisms, in addition to the

practical difficulty in obtaining access to biological fluids from sufficiently-large, well-matched early-stage patient cohorts.

Previous work reported by our group explored the feasibility of using MS-based metabolomics to detect ovarian cancer.[22, 26] While successful, these studies mainly involved late-stage (III/IV) OC patients in which metastasis had occurred, which are easily detectable because the disease is systemically widespread. The current work aims to build upon the previous studies by probing the metabolome to determine phenotypic fingerprints associated with early-stage OC.

3.3 Experimental Details

3.3.1 Chemicals

Healthy human blood serum (S7023-50 mL) was purchased from Sigma-Aldrich Corp. (St. Louis, MO, USA). FMOC-L-Proline was procured from Chem-Impex International, Inc. (Wood Dale, IL, USA). Leucine enkephalin was obtained from ERA (Golden, CO, USA). Arginyl-glycyl-aspartic acid, L-Fucose, L-Rhamnose, 1,5-Anhydrosorbitol, D-Fucose, L-Rhamnulose, 2-Deoxy-D-glucose, and 2-Deoxy-D-galactose were acquired from Sigma-Aldrich Corp. Bilirubin and suberic acid were obtained from Alfa Aesar (Ward Hill, MA, USA). Ricinoleic acid was purchased from MP Biomedicals (Santa Ana, CA, USA). Docosahexaenoic acid and 3-oxo stearic acid were acquired from Cayman Chemical Company (Ann Arbor, MI, USA). LysoPE(16:0) was obtained from Avanti Polar Lipids, Inc. (Alabaster, AL, USA). LC-MS grade methanol was purchased from J.T. Baker Avantor Performance Materials, Inc. (Center

Valley, PA, USA). Ultrapure water with 18.2 MΩ cm resistivity (Barnstead Nanopure, Thermo Fisher Scientific Inc., Waltham, MA, USA) was used to prepare chromatographic mobile phases.

3.3.2 *Dicer-Pten* Double-Knockout Mice

Dicer-Pten DKO (*Dicer*$^{flox/flox}$ *Pten*$^{flox/flox}$ *Amhr2*$^{cre/+}$) mice were generated by mating males (*Dicer*$^{flox/flox}$ *Pten*$^{flox/flox}$ *Amhr2*$^{cre/+}$) with females (*Dicer*$^{flox/flox}$ *Pten*$^{flox/flox}$). Female *Dicer*$^{flox/flox}$ *Pten*$^{flox/flox}$ (a genotype not carrying *Amhr2*$^{cre/+}$) mice were used as controls. Mice were housed in a vivarium with a controlled temperature of 21 °C. They were fed 5053 Irradiated PicoLab® Rodent Diet 20 and had access to drinking water supplied in bottles. *Dicer*$^{flox/flox}$ *Pten*$^{flox/flox}$ *Amhr2*$^{cre/+}$ DKO mice were sacrificed for this study in accordance to the animal protocol approved by the Institutional Animal Care and Use Committee (IACUC) at Baylor College of Medicine.

Blood samples were collected from 23 early-stage tumor (ET) and 10 late-stage tumor (LT) *Dicer-Pten* DKO mice (*Dicer*$^{flox/flox}$ *Pten*$^{flox/flox}$ *Amhr2*$^{cre/+}$) in addition to 21 control mice (*Dicer*$^{flox/flox}$ *Pten*$^{flox/flox}$). Murine blood samples were collected into serum separator tubes. Serum was obtained by centrifugation at 14,000 rpm for 5 min at room temperature. Immediately after centrifugation, 200 µL serum aliquots were frozen and stored at -80 °C until ultra performance liquid chromatography-mass spectrometry (UPLC-MS) analysis.

3.3.3 Human Cohort Description

Age-matched blood serum samples from 24 early-stage ovarian cancer patients (age range 40-84, mean age 58 ± 11 years) and 40 healthy women (age range: 40-84, mean age 57 ± 12 years) were acquired from the Ovarian Cancer Institute laboratory at the Georgia Institute of Technology. The healthy women population consisted of patients with histology considered within normal limits (WNL), despite the fact some had documented cysts. All donors were required to fast and to avoid medicine and alcohol for 12 hours prior to sampling, except for certain allowable medications (e.g., diabetics were allowed insulin). The mean ages of the 2 respective group populations were not significantly different (unpaired *t*-test, n = 64, $p = 0.67$). Blood samples were collected at Northside Hospital (Atlanta, GA) by venipuncture from each donor into evacuated blood collection tubes that contained no anticoagulant after approval by the Institutional Review Board (IRB). Serum was obtained by centrifugation at 5000 rpm for 5 min at 4 °C. Immediately after centrifugation, 200 µL aliquots of serum were frozen and stored at -80 °C for further use.

3.3.4 Serum Sample Preparation and Experimental Design

Commercially available healthy human blood serum was used to optimize the serum sample metabolite extraction protocol and UPLC-MS method. Serum samples were thawed on ice prior to sample preparation. Methanol was added to 100 µL of each serum sample in a 3:1 ratio to precipitate proteins. Samples were vortex-mixed for 10 s and centrifuged at 13,000 rpm for 7 min. After centrifugation, 350 µL of supernatant were transferred to new microcentrifuge tubes and, after the addition of 400 µL of

ultrapure water, frozen at -80 °C for 2 h. Subsequently, samples were lyophilized for 24 h at -50 °C and 25 mTorr using a VirTis bench top freeze-dryer (SP Industries, Stone Ridge, NY, USA). Sample residues were reconstituted in 100 µL of water/methanol (80:20 v/v, initial UPLC gradient conditions), and analyzed by UPLC-MS. Blank samples, consisting of ultrapure water, underwent the same process as serum samples. All samples were randomized prior to UPLC-MS analysis. Solvent and sample preparation blanks were jointly analyzed with serum samples. Quality control (QC) samples (15 µM FMOC-L-Proline and leucine enkephalin solution in ultrapure water) were analyzed every 5 hours to verify that retention time, peak shape and intensity were stable for the duration of the analysis. The relative standard deviation of the retention times, peak areas, and intensities of the monoisotopic ions obtained from extracted ion chromatograms were less than 15% over the duration of the experiments.

3.3.5 Metabolic Profiling *via* Ultra Performance Liquid Chromatography-Mass Spectrometry

UPLC-MS analysis was performed using a Waters ACQUITY UPLC H Class system fitted with a Waters ACQUITY UPLC BEH C_8 column (2.1 × 100 mm, 1.7 µm particle size) for *Dicer-Pten* DKO mice experiments or a Waters ACQUITY UPLC BEH C_{18} column (2.1 × 50 mm, 1.7 µm particle size) for the human cohort experiments, and coupled to a Xevo G2 QTOF mass spectrometer (Waters Corporation, Manchester, UK) with a typical resolving power of 25,000 M/Δ*m* FWHM and mass accuracy of 1.8 ppm at m/*z* 554.2615. The instrument was operated in negative ion mode with a probe capillary voltage of 2.5 kV, and a sampling cone voltage of 45 V. The ion source and desolvation

temperatures were 120 °C and 350 °C, respectively; the nitrogen desolvation flow rate was 800 L h^{-1}, and the cone desolvation flow rate was 50 L h^{-1}. The mass spectrometer was calibrated across the 50-1200 m/*z* range using a 0.5 mM sodium formate solution prepared in 90:10 2-propanol:water v/v. Data were mass corrected during acquisition using a leucine enkephalin reference spray (LockSpray) infused at 2 µL min^{-1}. Data were acquired in the 50-1200 m/*z* range and the scan time was set to 1 s. Data acquisition and processing was carried out using MassLynx v4.1. The chromatographic method for sample analysis involved elution with water (mobile phase A) and methanol (mobile phase B) at a flow rate of 0.40 mL min^{-1}. The following gradient program was used for analysis of *Dicer-Pten* DKO samples: 0-15 min 20-90% B; 15-19 min 90% B. The following gradient program was used for the analysis of human cohort samples: 0-15 min 20-90% B; 15-23 min 90% B. Both gradients were returned to their initial conditions over a period of 11 min after each sample injection. Column temperatures were set to 60 °C, the autosampler tray temperature was set to 5 °C, and the injection volume was 2 µL. Technical duplicates were acquired. UPLC-MS/MS experiments were performed by acquiring product ion mass spectra with applied voltages of 10, 20, and 30 V in the collision cell, using ultra high purity argon (≥ 99.999%) as the collision gas.

3.3.6 Data Analysis for Dicer-Pten DKO Mouse Model Study

Following UPLC-MS, spectral features (retention time (R_t), m/*z* pairs) were extracted from the data using MZmine 2.0 software.[31] This procedure involved chromatogram alignment, peak identification and integration, peak area extraction, and normalization after curation of the data matrix. The data matrix curation consisted of the

removal of signals that were present in the blank samples, the solvent, or were not present in at least 50% of the serum samples. The curated data matrix was utilized to build a model for sample class discrimination via oPLS-DA and to down-select a smaller panel of discriminant features through the usage of a genetic algorithm (MATLAB Version 7.13.0, The MathWorks, Inc., Natick, MA, USA with PLS_Toolbox v.6.71, Eigenvector Research, Inc., Wenatchee, WA, USA). A panel of 18 discriminant features had the lowest root-mean-square error of cross-validation (RMSECV) at the conclusion of the genetic algorithm variable selection process. The parameters for genetic algorithm variable selection were as follows: population size: 64, variable window width: 1, % initial terms (variables): 10, target minimum # of variables: 8, target maximum # of variables: 15, penalty slope: 0.05, maximum generations: 150, % at convergence: 79.7, mutation rate: 0.005, crossover: double, regression choice: PLS, # of latent variables: 6, cross-validation: random, # of splits: 5, # of iterations: 5, replicate runs: 20. PLS-DA models were orthogonalized and internally cross-validated using 10 iterations of random sample subsets with 5 data splits. Data were preprocessed by autoscaling prior to oPLS-DA analysis. Principal component analysis (PCA) was also performed to inspect data before and after genetic algorithm variable selection (i.e., on all of the extracted spectral features and only the discriminant feature panel).

3.3.7 Data Analysis for Human Cohort Study

Following UPLC-MS, spectral features (R_t, m/z pairs) were extracted from the data using MZmine 2.0 software.[31] This procedure involved chromatogram alignment, peak identification and peak area extraction, in addition to gap-filling and normalization

after curation of the data matrix. The data matrix curation consisted of the removal of signals that were present in the blank samples and the solvent. Also, signals that were not present in at least 50% of the healthy samples or 50% of samples belonging to either ovarian cancer histological subtype used in this study (i.e., papillary serous or endometrioid) were eliminated from the data matrix. Linear SVMs[32] in-house developed using LIBLINEAR[33] were utilized to build multivariate sample classification models and to find the minimum set of discriminant features needed to differentiate the early-stage OC patient samples from those of the healthy women with the highest accuracy, specificity, and sensitivity. SVMs are effective at handling high dimensionality data as those produced in the present work and have been widely applied to metabolomics studies.[22, 26, 34-36] Moreover, they are ideal for limited-size datasets since the risk of data overfitting is reduced.[37] All data were preprocessed by autoscaling before analysis.

For SVM analysis, linearly-separable samples represented as a row vector $\boldsymbol{x}$, had membership of two classes g (= H or D), where H stands for healthy women and D for OC disease with labels c = -1 for class H, and +1 for class D. All of the samples were used to build a classification model which was internally validated using leave-one-out cross-validation. The decision function that separated the two classes, defined here as the "OC detection score", was as follows:

$$OC\ detection\ score = b + \sum_{j=1}^{J} w_j x_{ij} \qquad [1]$$

$$g(\boldsymbol{x}_i) = sgn(\boldsymbol{w}\boldsymbol{x}_i' + b) = sgn(OC\ detection\ score) \qquad [2]$$

where *w* and *b* are the weight and bias parameters that were determined from the SVM classification model and J is the total number of features. The sign of the OC detection score determined which class a sample was assigned to: class H if negative and class D if positive. In this classification function, the two classes were divided in the dataspace by a hyperplane $\boldsymbol{wx}' + b = 0$ that maximized the margins between samples of different classes. The margin between the two classes was defined such that:

$$\boldsymbol{wx}' + b \geq 1, \qquad c = +1 \qquad [3]$$

$$\boldsymbol{wx}' + b \leq -1, \qquad c = -1 \qquad [4]$$

A recursive feature elimination (RFE) method,[22] structured in the SVM analysis process as a nested leave-one (sample)-out design, was used to find the minimum set of discriminant features that maximized the accuracy of the SVM classification model. The design consisted of two loops. The outer loop used one randomly selected sample as a test sample while the remaining samples constituted the inner loop training set. In the inner loop, one sample was iteratively left out while the remaining samples were used as a training set to calculate weights for each metabolic feature and predict the left-out sample. These weights were averaged over all inner loop iterations, and the resultant SVM model was used to classify the outer loop test sample. Afterwards, a different outer loop test sample was selected and classified after another set of inner loop iterations were completed. This process ensued until all samples had been used as outer loop test samples. Once the averaged weights resulting from each inner loop process were summed, the feature that ranked least important (i.e., had the lowest summed weight) was discarded from the remaining feature set ($w_i = 0$), and the inner/outer loop processes

began again. The aforementioned procedure was repeated until all metabolic features were ranked. The performance characteristics for the RFE-SVM method were generated based on the percentage of test samples accurately classified during each phase of the feature elimination procedure. From this process, a feature panel with high accuracy, sensitivity, and specificity was chosen and tested again across all samples.

PLS-DA was also performed to inspect data before and after discriminant feature selection via RFE. PLS-DA models were orthogonalized (oPLS-DA) and internally cross-validated using 5 iterations of random sample subsets with 8 data splits. Data were preprocessed by autoscaling prior to oPLS-DA analysis.

3.3.8 Discriminant Feature Identification Procedure

Metabolite identification was attempted for discriminant features. Mass spectral ion adduct analysis was first performed to ensure the unambiguous assignment of the signal of interest in each mass spectrum. The adduct ions that were investigated in the mass spectra included $[M - H]^-$, $[M + Cl]^-$, $[M + CH_3COO]^-$, $[M + HCOO]^-$, $[M + Na - 2H]^-$, $[M + K - 2H]^-$, $[M - H_2O - H]^-$, $[M + H_2O - H]^-$, and $[2M – H]^-$ species, which are usually observed in negative electrospray ionization mode. The theoretical m/z values for these species were calculated and compared with the experimental values from mass spectral signals. For spectra in which multiple adducts were not present, the accurate mass of the candidate neutral molecule was calculated based on the assumption that the m/z value observed corresponded to the $[M - H]^-$ ionic species. For mass spectra in which multiple adducts were present, the $[M - H]^-$ spectral signal was determined and the accurate mass of the metabolic candidate neutral molecule was calculated based on it.

Elemental formulae were generated based on the exact mass (maximum mass error of 10 mDa) and isotopic patterns of the features using MassLynx 4.1. The elements included in the formulae were constrained to C, H, N, O, P, and S. The lists of generated elemental formulae were searched against the METLIN database,[38] the LIPID Metabolites and Pathways Strategy (LIPID MAPS) database,[39] and the human metabolome database (HMDB).[40] MetaboSearch[41] was also utilized to search the aforementioned databases solely using neutral masses with a mass accuracy of 20 ppm. Tandem MS data could not be acquired for discriminant features where the precursor ion abundance was not high enough for sensitive quadrupole selection and MS/MS due to ion transmission losses.[42] The MS/MS METLIN database, MassBank,[43] and literature searches were used to further confirm the identity of the candidates for which MS/MS data was successfully acquired. Additionally, fragmentation patterns were manually analyzed in a few cases to discriminate between different isobaric species. For the *Dicer-Pten* DKO mouse model study, available chemical standards were purchased to validate tentative metabolite identities by chromatographic retention time matching and/or MS/MS fragmentation pattern matching. These chemical standards also served the purpose of eliminating possible metabolite candidates from the tentative identification list.

3.4 Metabolomic Serum Profiling Detects Early-Stage High-Grade Serous Ovarian Cancer in a Mouse Model

3.4.1 DKO Mouse Cohort

Typically, between 4 and 7 months, DKO mice develop HGSCs of the fallopian tubes, which later spread to envelop the ovaries and at a late stage metastasize throughout the abdominal cavity including peritoneal tissues (Figure 3.1). For the purpose of this study, early-stage tumors in DKO mice were defined as those confined to the fallopian tube without any sign of metastasis. Eventually, all DKO mice die from metastatic HGSCs after developing hemorrhagic ascites, an accumulation of hemorrhagic fluid in the abdominal cavity. The typical life span of these mice is 6.5–13 months.[16]

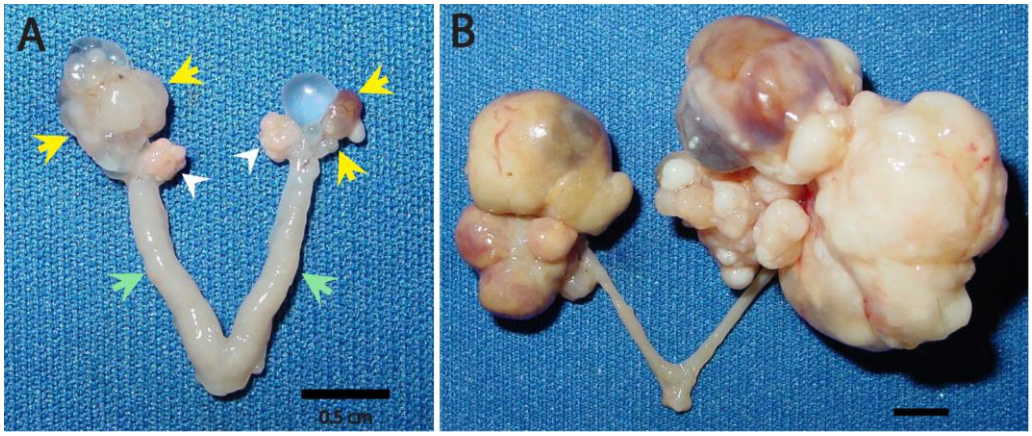

Figure 3.1: Early- and late-stage high-grade serous carcinomas (HGSCs) in DKO mice (*Dicer* $^{flox/flox}$ *Pten* $^{flox/flox}$ *Amhr2* $^{cre/+}$). (A): Early fallopian tube tumors (yellow arrows) formed in a 6.8-month-old DKO mouse used in this study with normal ovaries (white arrowheads) and uterus (green arrows). (B): Massive fallopian tube tumors that engulfed the ovaries in a 10.7-month-old DKO mouse with late-stage HGSCs.

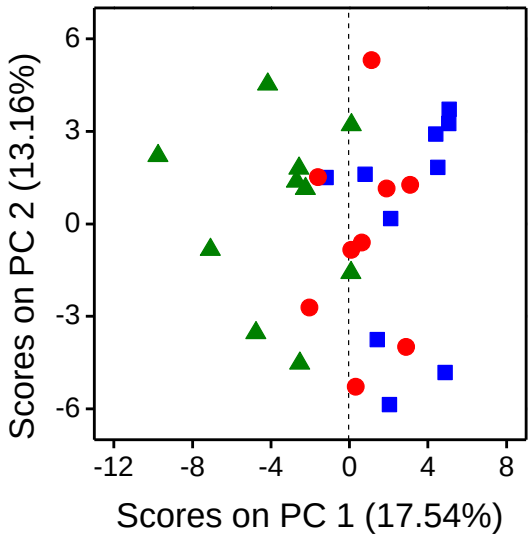

Figure 3.2: Principal Component Analysis (PCA) scores plot of early-stage tumor *Dicer-Pten* DKO (red circles), late-stage tumor *Dicer-Pten* DKO mice (green triangles), and control mice (blue squares). The model consisted of 2 PCs with 30.70% total captured variance.

Initially, an exploratory experiment was conducted to determine to what extent UPLC-MS metabolic profiling could differentiate the blood sera of 9 ET (mean age 319 ± 36 days) and 10 LT (mean age 309 ± 34 days) DKO mice, in addition to 10 control (mean age 342 ± 80 days) mice. For this experiment, DKO mice were staged based on the presence or lack of ascites—LT DKO mice had developed ascites while ET DKO mice had not. Unsupervised PCA of the resultant data showed clear separation of LT DKO mice from ET DKO mice and control mice along the 1st principal component (unpaired t-test, n = 29, p = 0.002) (Figure 3.2). However, both ET DKO mice and control mice were clustered together. As expected, this result somewhat reflects what is clinically observed

for the diagnosis of human ovarian carcinomas, i.e. late-stages in which metastasis has occurred are more easily detectable because the disease is systemically widespread. Yet, detection of early-stages is challenging since the disease is still localized and asymptomatic.

At this stage the focus of the study was shifted to the detection of early-stage HGSC in DKO mice by comparing metabolomic profiles between DKO mice with early-stage HGSC against control mice, as this is the equivalent of the clinically-relevant challenge for human HGSC detection. Therefore, a set of 14 *Dicer-Pten* DKO mice with early-stage HGSC (mean age 206 ± 19 days) and 11 control mice (mean age 211 ± 30 days) was investigated via an alternative approach involving supervised multivariate analysis. To ensure complete confidence in the class membership of each mouse in the cohort, ET DKO mice were sacrificed after blood sample collection to confirm the early-tumor status by ensuring that all had primary tumors located on their fallopian tubes with no visible metastasis. Table 3.1 describes detailed information of the mice included in these experiments as well as the tumor status verified for each ET DKO mouse.

Table 3.1: Early-Stage Tumor *Dicer-Pten* Double Knockout Mice and Control Mice Characteristics.

Early-Stage Tumor DKO Mice					Control Mice			
ID	**Date of Birth**	**Age**		**Tumor Status**	**ID**	**Date of Birth**	**Age**	
		(days)	**(months)**				**(days)**	**(months)**
310	5/8/12	181	6.0	Tumor on each oviduct	109	2/16/12	257	8.6
312	5/8/12	181	6.0	Tumor on each oviduct	115	2/28/12	245	8.2
315	5/8/12	181	6.0	Tumor on each oviduct	116	2/28/12	245	8.2
252	4/16/12	199	6.6	Tumor on each oviduct	156	3/14/12	230	7.7
255	4/16/12	199	6.6	Tumor on one oviduct	222	5/12/12	170	5.7
215	4/10/12	204	6.8	Tumor on one oviduct	223	4/10/12	202	6.7
216	4/10/12	205	6.8	Tumor on one oviduct	218	4/3/12	209	7.0
217	4/10/12	205	6.8	Tumor on one oviduct	309	4/3/12	209	7.0
219	4/10/12	205	6.8	Tumor on each oviduct	311	4/3/12	209	7.0
205	4/8/12	206	6.9	Precursor tumor lesion on each oviduct	372	5/8/12	175	5.8
220	4/3/12	211	7.0	Precursor tumor lesion on each oviduct	453	5/8/12	175	5.8
256	4/16/12	217	7.2	Tumor on each oviduct				
143	3/10/12	235	7.8	Precursor tumor lesions on each oviduct				
153	3/14/12	250	8.3	Tumor on each oviduct				

3.4.2 Multivariate Classification Performance

MZmine data processing extracted 934 features (R_t, m/*z* pairs) from the UPLC-MS data from both control and ET DKO mice. These extracted features were utilized to build an oPLS-DA model which classified the respective serum samples from each class. Performance characteristics of the initial oPLS-DA analysis of the data matrix that included all 934 metabolic features (Figure 3.3a and 3.3b) were 76%, 68%, and 83% for the cross-validated accuracy, sensitivity, and specificity, respectively. A total of five murine serum samples were misclassified. This 3 latent variable model interpreted 35.35% and 93.64% variance from the X- (feature peak areas) and Y- (mouse class membership) blocks, respectively. Though the model performance was not entirely poor, genetic algorithms were used to attain a smaller, but more robust, metabolic feature set that could serve to better discriminate between control and ET DKO mice with higher cross-validated accuracy, sensitivity, and specificity. A panel of 18 metabolic features with the lowest RMSECV was selected through the genetic algorithm variable selection process. oPLS-DA modeling with this smaller panel (Figure 3.3c and 3.3d) resulted in 100% cross-validated accuracy, sensitivity, and specificity; therefore, no mice were misclassified. This model interpreted 33.12% and 98.30% of the X- and Y-block variances, respectively, with only 2 latent variables —one less than the model using all 934 metabolic features. Furthermore, the captured Y-block variance was slightly higher, thereby, demonstrating that the down-selected panel of 18 features is more informative than the initial set.

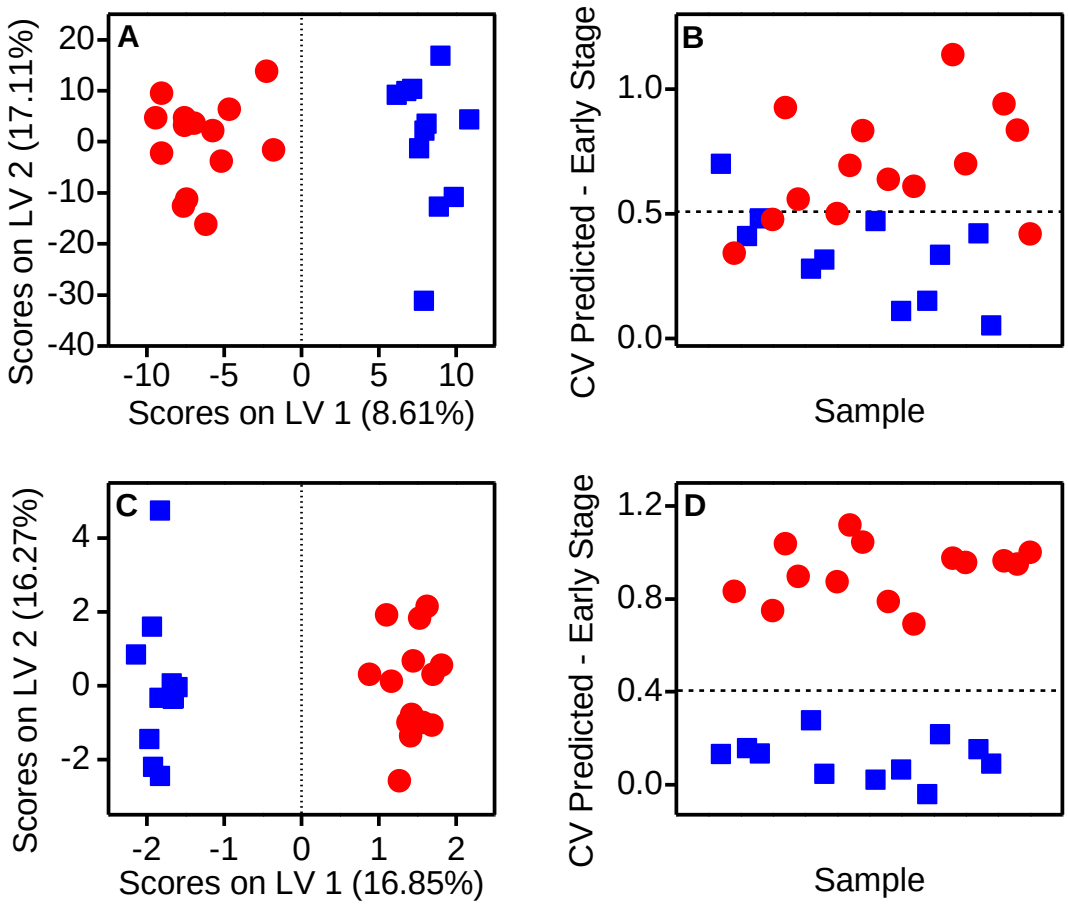

Figure 3.3: Orthogonal projection to latent structures-discriminant analysis (oPLS-DA) models of early-stage tumor (ET) *Dicer-Pten* DKO (red circles) vs. control mice (blue squares). (A): oPLS-DA calibration scores plot using the total initial set of 934 spectral features. The model consisted of 3 LVs with 35.35% and 93.64% total captured X- and Y-block variances, respectively. The cross-validated accuracy, sensitivity, and specificity were 76%, 68%, and 83%, respectively. (B): The corresponding ET cross-validated prediction plot for (A). There were 5 misclassified mice. (C): oPLS-DA calibration scores plot using the 18 discriminant metabolic feature panel obtained from genetic algorithm variable selection. The model consisted of 2 LVs with 33.12% and 98.30% total captured X- and Y-block variances, respectively. The accuracy, sensitivity, and specificity were all 100%. (D): The corresponding ET cross-validated prediction plot for (C). There were no misclassified mice.

Of the 18 selected features, the concentration levels of 9 metabolites were found to increase and 9 metabolites were found to decrease in ET DKO mice. While only 11 of these concentration changes were univariately significant (Mann-Whitney U test, n = 25, $p \leq 0.05$; Figure 3.4), the co-varying concentrations of all 18 metabolic features allowed us to distinguish the detected metabolomes of control and ET DKO mice in multivariate space; accordingly, they all display discriminatory power when collectively included as part of a joint panel. The concentration levels of those features that were not statistically different between ET DKO and control mice in a univariate fashion could also be a result of the relatively modest sample size used in this study.

To further investigate these results, PCA was utilized to evaluate the performance of the 18-feature discriminant panel in an unsupervised manner. Scores plots were generated for both the initial set of 934 metabolic features and the 18 discriminant feature panel (Figure 3.5). Using the initial set, 3 principal components interpreting 46.43 % of the data matrix variance showed no distinct grouping of the sample classes (Figure 3.5a). However, better clustering was observed with the 18-feature 3-principal component PCA model which interpreted 53.49% of the variance (Figure 3.5b). The PCA class separation was statistically significant and occurred along the 1st principal component (unpaired *t*-test, n = 25, $p = 0.0055$), providing further evidence of the robustness of the 18-feature discriminant panel.

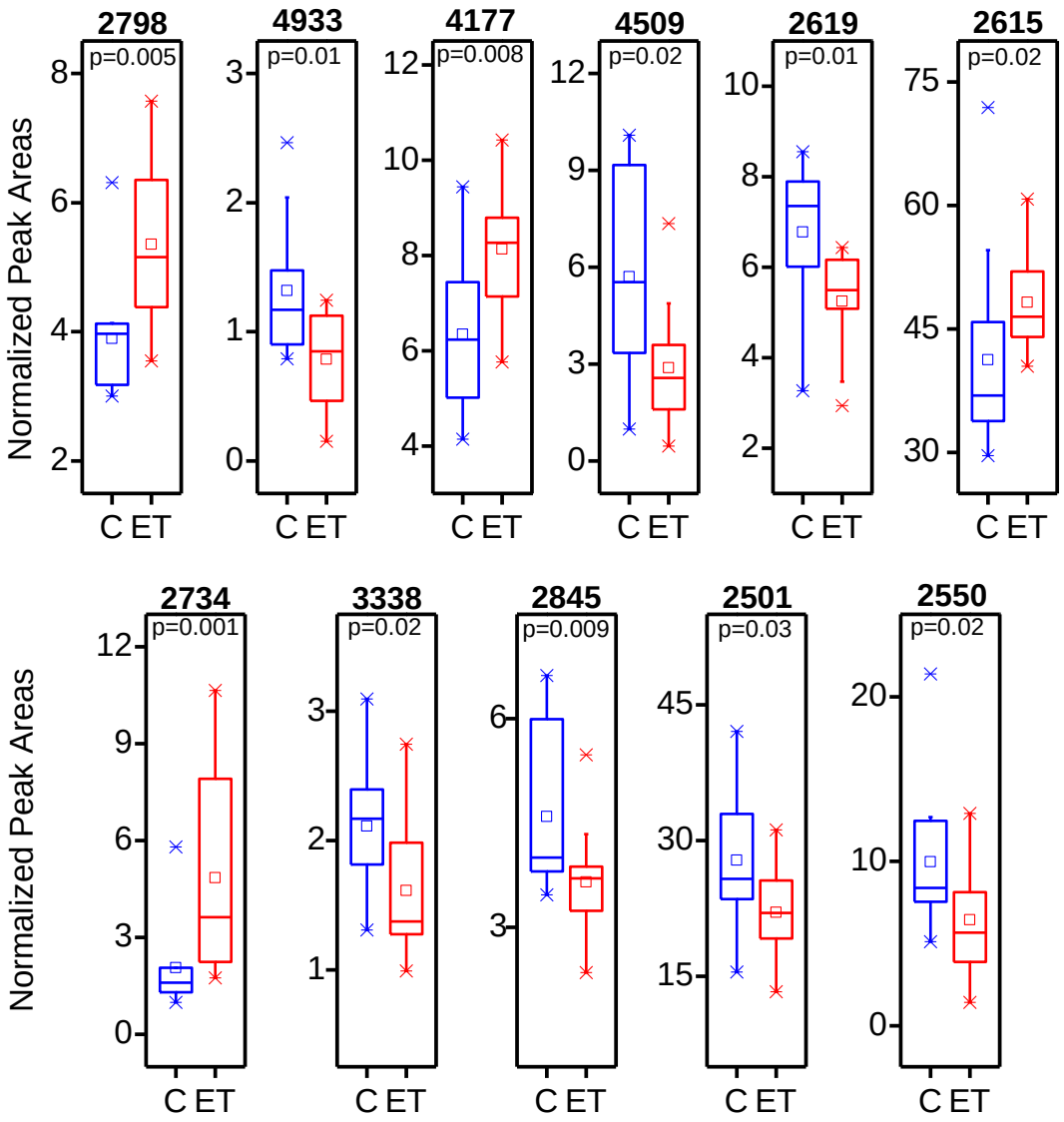

Figure 3.4: Discriminatory spectral features having statistically significant univariate changes between early-stage tumor (ET) *Dicer-Pten* DKO mice (n = 14) and control (C) mice (n = 11). *P*-values were calculated using the Mann-Whitney U test. Box plots with mean (square), median, upper and lower quartile, outliers, and minimum and maximum (whiskers) data values are displayed. Feature ID numbers are indicated on top of each case.

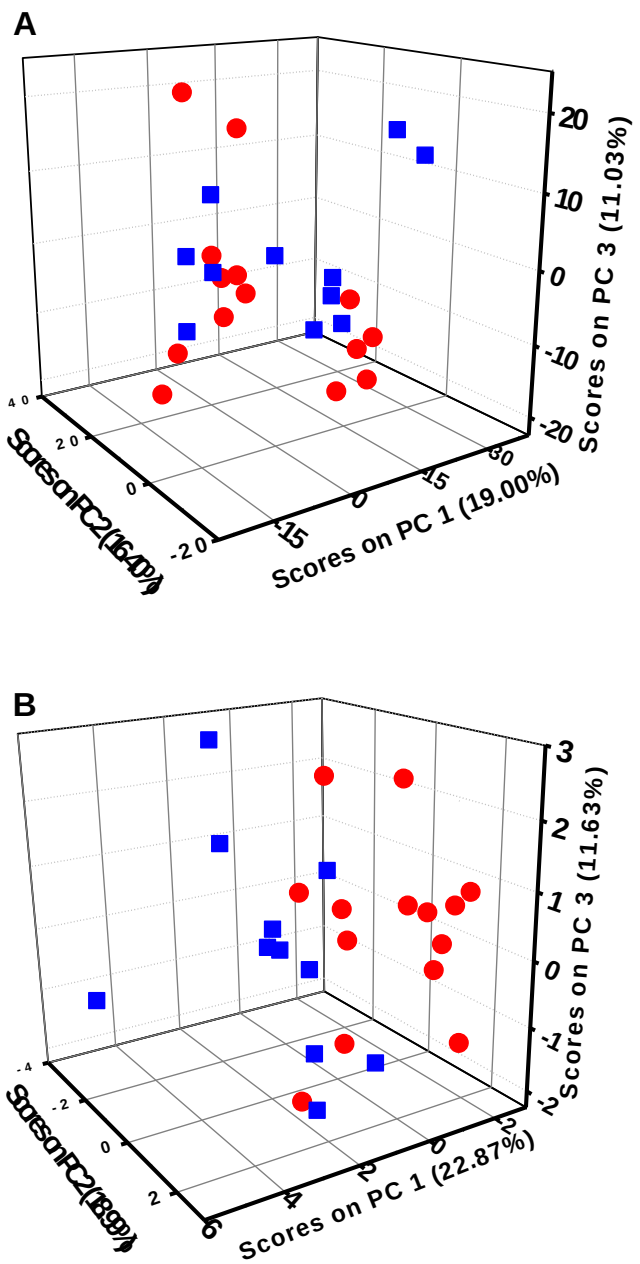

Figure 3.5: Principal Component Analysis (PCA) of early-stage tumor *Dicer-Pten* DKO (red circles) and control mice (blue squares). (A): PCA scores plot using the initial set of 934 spectral features. The model consisted of 3 PCs with 46.43% total captured variance. (B): PCA scores plot using the 18 discriminant feature panel obtained from genetic algorithm variable selection. The model consisted of 3 PCs with 53.49% total captured variance.

3.4.3 Discriminant Metabolite Identification

Metabolite identification was attempted for the 18 discriminant feature subset once PCA confirmed the robustness of the oPLS-DA model. The metabolite identification procedure is illustrated in Figure 3.6 for feature 2501. Typical total ion UPLC-MS chromatograms for control and ET DKO mice are shown in Figure 3.6a and 3.6b, respectively. The apparent similarity of these chromatograms hints that subtle metabolic differences occur as HGSCs initially form and begin to proliferate and demonstrates the need for multivariate methods to extract these differences from the detected metabolomes. Highly-selective accurate mass extracted ion chromatograms for each discriminant feature (Figure 3.6c) were generated thanks to the high resolving power of the time-of-flight mass analyzer utilized. The corresponding averaged mass spectra were used for adduct ion analysis (Figure 3.6d) to ensure that the correct exact mass was assigned to the spectral features of interest. The presence of multiple adducts for some features further assisted in chemical assignments. The ensuing isotopic patterns and accurate masses were used to generate lists of possible candidate elemental formulae that were searched against metabolite databases. Furthermore, UPLC-MS/MS experiments were conducted to obtain fragmentation patterns that helped confirm the metabolite identities of the 18 feature subset. Tandem MS spectra were compared to those in metabolite databases and/or scientific literature and manually analyzed as well (Figure 3.6e). Lastly, standards of the commercially-available metabolites were analyzed by UPLC-MS and UPLC-MS/MS to further verify the identity of the candidates by retention time and mass spectral matching, as well as eliminating non-matches.

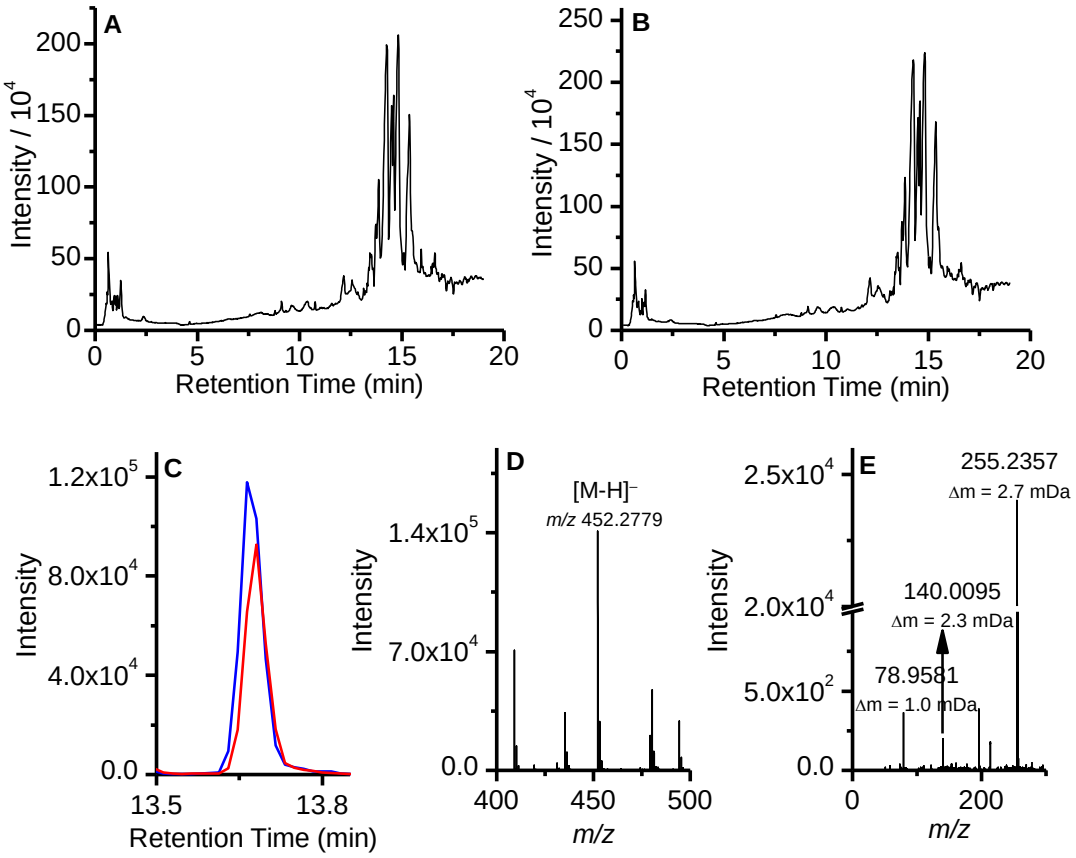

Figure 3.6: Typical total ion chromatograms obtained for serum samples from (A) a control mouse and (B) a *Dicer-Pten* early-stage tumor DKO mouse. (C): Extracted ion chromatogram for m/*z* 452.2779 ± 0.0050 generated from a control mouse sample (blue line) and a *Dicer-Pten* early-stage tumor DKO mouse (red line). These were generated from the data shown in A and B, respectively. (D): Mass spectrum for the discriminant metabolic feature with m/*z* 452.2779. (E): Tandem MS spectrum for the m/*z* 452.2779 precursor ion using a collision cell voltage of 30 V. The matching of the experimental tandem MS fragmentation pattern with the metabolite candidate is illustrated by the mass errors calculated as differences with the values in the MassBank database, in addition to manual fragment analysis.

Eleven of the 18 metabolic features were identified by MS and MS/MS, while 4 were further confirmed chromatographically by chemical standards (Table 3.2). These 4 were ricinoleic acid (Figure 3.4; +0.23 fold change), bilirubin (Figure 3.4; -0.74 fold change), suberic acid (-0.42 fold change) and LysoPE(16:0) (Figure 3.4; -0.33 fold change). Additional identified features (Table 3.2) include di- and triglycerides, phospholipids, bile acids, and terpenes. Four of the features in Table 3.2 were tentatively matched to exogenous metabolites which have no known biofunctions, such as Peltatol A (lignan), and/or metabolites yet to be reported in mammals, such as (4E,8E,10E-d18:3)sphingosine (sphingoid base). These tentative metabolite identifications are listed in Table 3.3, and the metabolite chemical class is still listed in Table 3.2 (italicized).

Table 3.2: Confirmed and Tentative Metabolites Identified as Discriminatory Between Early-Stage Tumor *Dicer-Pten* Double Knockout Mice and Control Mice. Metabolites confirmed by retention time matching with commercially-available standards are highlighted in bold font.

Feat. Code	Avg. Ret. Time[1] (min)	Avg. m/z[1]	Fold Change[2]	P value[3]	Ion Type	Elemental Formula	Theoretical m/z	Mass Error (mDa)	Tentative Metabolite Identification	Ref.
2734	11.88	294.2410	1.23	0.0011	$[M-H]^-$	$C_{18}H_{33}NO_2$	294.2439	2.9	*Sphingoid base*	18, 21
2798	15.53	363.2883	0.46	0.0051	$[M-H]^-$	$C_{23}H_{40}O_3$	363.2905	2.2	24-Nor-5β-cholane-3α,6α,23-triol 24-Nor-5β-cholane-3α,7α,12α-triol 24-Nor-5β-cholane-3α,7α,23-triol 24-Nor-5β-cholane-3α,7β,23-triol 24-Nor-5β-cholane-3α,12α,23-triol 4-Hydroxy-3-(16-methylheptadecyl)-2H-pyran-2-one 2-(8-[3]-ladderane-octanyl)-sn-glycerol	17, 22, 25
4177	14.81	889.7234	0.36	0.0075	$[M-H]^-$	$C_{58}H_{98}O_6$	889.7290	5.6	TG(55:7)	44
4466	13.73	625.4240	0.26	0.73	$[M-H]^-$	$C_{42}H_{58}O_4$	625.4262	2.2	*Lignan*	45
2615	**12.60**	**297.2414**	**0.23**	**0.015**	**$[M-H]^-$**	**$C_{18}H_{34}O_3$**	**297.2435**	**2.1**	**Ricinoleic acid**	**27**
2489	13.37	476.2772	0.22	0.27	$[M-H]^-$	$C_{23}H_{44}NO_7P$	476.2783	1.1	LysoPE(18:2)	21, 27, 46
3302	11.76	163.0578	0.11	0.40	$[M-H]^-$	-	-	-	N/A	-

Table 3.2 (continued).

3154	17.31	780.5530	0.061	0.77	$[M-H]^-$	$C_{44}H_{80}NO_8P$	780.5550	1.9	PE(39:4)	22, 27, 47
2537	13.33	219.1724	0.0059	0.40	$[M-H]^-$	$C_{15}H_{24}O$	219.1750	3.0	*Terpene derivative*	48
4509	16.85	627.5034	-0.99	0.018	$[M-H]^-$	$C_{40}H_{68}O_5$	627.4994	4.0	DG(37:5)	49
4933	**11.49**	**583.2555**	**-0.74**	**0.011**	**$[M-H]^-$**	**$C_{33}H_{36}N_4O_6$**	**583.2557**	**0.7**	**Bilirubin**	**25**
2550	14.70	711.3057	-0.63	0.025	$[M-H]^-$	-	-	-	N/A	-
4590	1.23	345.1557	-0.52	0.29	$[M-H]^-$	$C_{16}H_{26}O_8$	345.1555	0.2	Uroterpenol-O-glucuronide	50
						$C_{17}H_{22}N_4O_4$	345.1568	1.1	Ala-Trp-Ala[4]	22
						$C_{14}H_{26}N_4O_4S_1$	345.1602	4.5	Lys-Cys-Pro[4]	22
3226	**1.22**	**173.0810**	**-0.42**	**0.43**	**$[M-H]^-$**	**$C_8H_{14}O_4$**	**173.0820**	**0.9**	**Suberic acid**	**27, 51**
2619	13.43	327.2295	-0.37	0.013	$[M-H]^-$	$C_{22}H_{32}O_2$	327.2330	3.5	*Fatty acid, terpene, or other phenol derivative*	27, 48
2501	**13.68**	**452.2783**	**-0.33**	**0.033**	**$[M-H]^-$**	**$C_{21}H_{44}NO_7P$**	**452.2783**	**2.1**	**LysoPE(16:0)**	**21, 27, 46**
3338	13.33	903.6231	-0.33	0.021	$[M-H]^-$	$C_{49}H_{93}O_{12}P$	903.6332	10.1	PI(O-18:0/22:2) PI(P-18:0/22:1) PI(O-20:0/20:2) PI(P-20:0/20:1)	27, 52
2845	0.46	402.7997	-0.33	0.0090	$[M-H]^-$	-	-	-	N/A	-

[1]These reported values were obtained after data processing with MZmine 2.0 software which generates average values calculated using all samples.
[2]Fold change was calculated as the base 2 logarithm of the average peak area ratios for ET DKO mouse samples and control samples.
[3]P values calculated using Mann-Whitney U test.
[4]Other polypeptide isomers are not listed; tandem MS data could not be acquired to confirm true isomer identity as their ion abundances were not high enough for quadrupole selection.

Table 3.3: Exogenous Metabolites or Metabolites Not Reported in Mammals Identified as Discriminatory Between Early-Stage Tumor *Dicer-Pten* DKO and Control Mice.

Feat. Code	Avg. Ret. Time[1] (min)	Avg. m/z[1]	Ion Type	Elemental Formula	Theoretical m/z	Δm (mDa)	Tentative Metabolite Identification *(Chemical class)*	Ref.
2734	11.88	294.2410	$[M-H]^-$	$C_{18}H_{33}NO_2$	294.2439	2.9	*(Sphingoid base)* (4E,8E,10E-d18:3)sphingosine	18, 21
4466	13.73	625.4240	$[M-H]^-$	$C_{42}H_{58}O_4$	625.4262	2.2	*(Lignan)* Peltatol A Peltatol B Peltatol C	45
2619	13.43	327.2295	$[M-H]^-$	$C_{22}H_{32}O_2$	327.2330	3.5	*(Fatty acid, terpene, or phenol derivative)*[3] 3,9,15-Docosatriynoic acid 8,11,14-Docosatriynoic acid 10,13,16-Docosatriynoic acid Neogrifolin Grifolin (Z,Z)-2-Methyl-5-(8,11,14-pentadecatrienyl)-1,3-benzenediol	27, 48

Table 3.3 (continued).

							(Terpene derivative) (2R,6S,7S,10Z)-beta-Santala-3(15),10-dien-12-ol alpha-Valerenol Isospathulenol alpha-Santal-10-en-12-ol alpha-Cyperol beta-Costol Isocyperol (R)-2,7(14),9-Bisabolatrien-11-ol (3S,6E)-6-Caryophyllen-15-al alpha-Bergamotenol Spathulenol beta-Santalol trans-beta-Santalol	
2537	13.33	219.1724	$[M-H]^-$	$C_{15}H_{24}O$	219.1750	3.0	beta-Betulenol 8alpha-3-Copaen-8-ol 4(15)-Copaen-11-ol 11-Copaen-4-ol Humulenol I Humulenol II Bisacurol Solavetivol Pentalen-13-ol Germacra-1(10),4,11(13)-trien-12-ol (+)-Khusimol albaflavenol (5R)-albaflavenol (5S)-albaflavenol 3-Copaene-2-ol Nootkatol	[48]

Table 3.3 (continued).
[1]These reported values were obtained after data processing with MZmine 2.0 software which generates average values calculated using all samples.

3.4.4 HGSC-related Metabolic Alterations

The identified discriminant metabolites cover a broad range of biomolecule classes and pathways (Figure 3.7), many of which have been individually reported to be involved in ovarian cancer proliferation. Table 3.2 summarizes these findings, which are discussed based on pathway/chemical class below. Although fold changes are reported in this discussion, the only statistically significant metabolite fold changes were those shown in Figure 3.4.

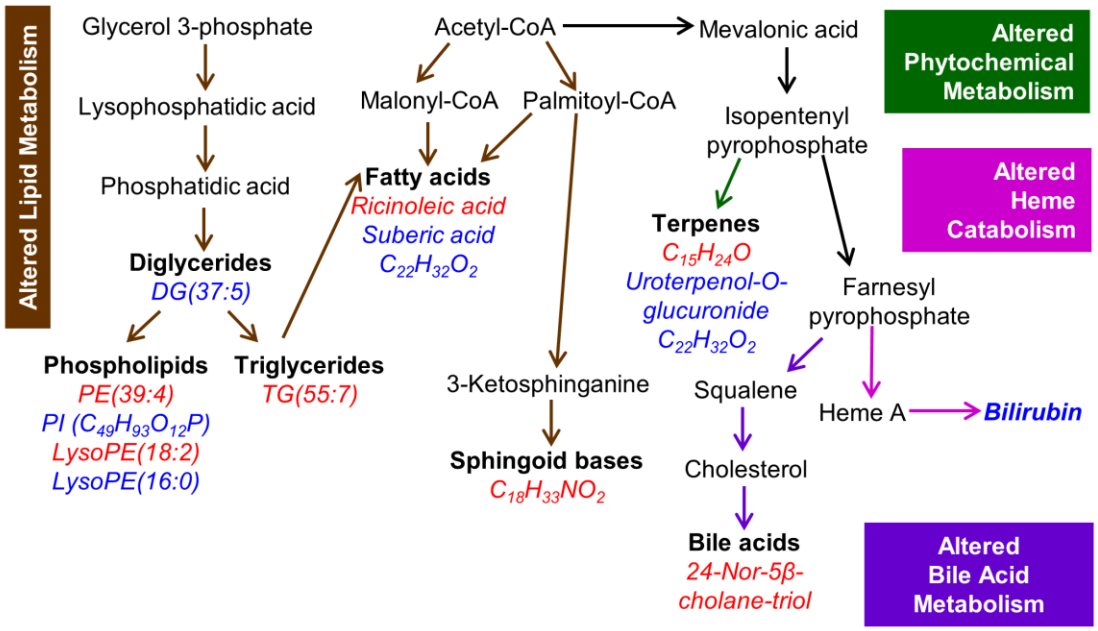

Figure 3.7: Overview of key altered metabolic/catabolic pathways observed in early-stage tumor (ET) *Dicer-Pten* DKO mice. Arrows are colored based on their corresponding altered pathways and represent both direct and indirect relationships between metabolites/metabolite classes. Metabolites from the 18 discriminant feature subset are italicized, and their corresponding metabolite class is in bold. Metabolites in red text have increased levels in ET DKO mice, while those in blue text have decreased levels in ET DKO mice. Metabolite relationships were derived from Kyoto Encyclopedia of Genes and Genomes,[53] MetaCyc,[54] and existing scientific literature.

3.4.4.1 Fatty Acids and Derivatives

The development of hormone-related cancers, such as ovarian cancer, may be influenced by fatty acids.[55] Fatty acids are involved in tumor cell signaling and growth, and are also used for energy expenditure, bulk membrane synthesis, and membrane-targeted protein modifications.[56, 57] Thus, it is not surprising that changes in fatty acid concentrations aid in detecting the early stages of HGSC development in the DKO mouse model. Five features in our discriminant subset were tentatively identified as fatty acids or their derivatives (Table 3.2; Figure 3.7): TG(55:7), ricinoleic acid, DG(37:5), suberic acid, and an unidentified possible fatty acid ($C_{22}H_{32}O_2$).

Estrogens, which play a major role in ovarian cancer proliferation,[58] have been linked with increased hepatic triglyceride production and secretion into blood in mice,[59] chickens,[60, 61] and humans.[62, 63] Accordingly, increased blood serum triglyceride concentrations have been linked to gynecological (ovarian, endometrial, cervical) cancer risk,[44] in agreement with our finding in ET DKO mice (Table 3.2; Figure 3.4; fold change: +0.36). Gercel-Taylor *et al.* found increased diglyceride levels in the ascites of late-stage ovarian cancer patients,[49] which is opposite to the decrease found in this study in the serum from mice with early-stage tumors (Table 3.2; Figure 3.4; fold change: -0.99). Since diglycerides are precursors to triglycerides, decreased diglyceride levels could be the result of increased triglyceride synthesis in the ET DKO mice, further demonstrating that HGSCs impact energy storage and generation. As a note, the identified odd-chain di- and triglycerides could also result from dietary intake, thus, hinting that fatty acid catabolism could be occurring at a decreased rate in ET DKO mice.

Suberic acid, a metabolic breakdown product of oleic acid, had decreased serum levels (Table 3.2; fold change: -0.42) in ET DKO mice. Menendez *et al.* found that oleic acid suppresses the transcriptional activity of the Her-2/*neu* oncogene in the SK-OV3 ovarian cancer cell line by up-regulating polyomavirus enhancer activator 3 (PEA3), a Her-2/*neu* promoter transcriptional repressor.[51] The detected decrease of suberic acid may thus be indirectly connected to the anti-cancer properties offered by oleic acid. Additionally, conjugated linoleic acid has been shown to suppress epidermal and mammary carcinogenesis in mice.[64, 65] The detected increase in serum levels of ricinoleic acid (Table 3.2; Figure 3.4; fold change: +0.23) in ET DKO mice could thus be related to its increased production for further conversion into other fatty acids such as conjugated linoleic acid, as reported for bacteria.[66, 67] In a metabolomics study, Denkert *et al.* detected decreased levels of free fatty acids (nonadecanoic acid, stearic acid, heptadecanoic acid) in invasive ovarian carcinoma tumors compared to borderline ovarian tumors.[27] Correspondingly, the serum levels of the unidentified possible $C_{22}H_{32}O_2$ fatty acid (Table 3.2; Figure 3.4; fold change: -0.37) were also decreased in the ET DKO mice.

3.4.4.2 Phospholipids

Cancer cells need a continuous supply of phospholipids for generation and maintenance of membrane integrity as well as for protein modifications.[68] Four features in our discriminant subset were tentatively identified as phospholipids (Table 3.2; Figure 3.7): LysoPE (18:2), PE(39:4), LysoPE(16:0), and $C_{49}H_{93}O_{12}P$ [PI(O-18:0/22:2), PI(P-18:0/22:1), PI(O-20:0/20:2), or PI(P-20:0/20:1)]. The detected changes in phospholipids suggest membrane degradation or morphological changes in HGSC-affected cells since

these lipids play a role in maintaining cellular membrane integrity. However, the interpretation of both increased and decreased levels is not straightforward, and could be the result of membrane instability and the subsequent attempt of cancer cells to repair them.

The PIK3CA oncogene, whose encoded protein phosphorylates phosphatidylinositols, as well as increased PI3-kinase activity have been implicated in ovarian carcinogenesis.[52] Links between PI3-kinase activation and increased cell survival due to the suppression of apoptosis have also been established.[69, 70] Moreover, increased PI3-kinase activity may aid in cancer proliferation by increasing cellular motility.[52] The observed PI ($C_{49}H_{93}O_{12}P$) decrease (Table 3.2; Figure 3.4; fold change: -0.33) in ET DKO mice could thus be attributed to increased PI3-kinase activity. As the conditional deletion of *Pten* from ET DKO mice activates the PI3-kinase signaling pathway, increased PI3-kinase activity could also be an artefact of the genetic manipulation of the mice.[16]

Serum levels of PE(39:4) (Table 3.2; fold change: +0.061) were increased in ET DKO mice. Previous work by our group also found that a collection of phosphatidylethanolamines aided in combined early- and late-stage HGSC detection in human subjects.[22] Moreover, choline kinase, an enzyme needed for the synthesis of phosphatidylethanolamines and phosphatidylcholines, has been shown to be increased in the EOC cell lines SK-OV3 and OVCAR-3.[47] Lysophosphatidylethanolamine has been shown to increase intracellular calcium concentrations and stimulate an unknown membrane receptor, causing chemotactic migration and cell invasion in the SK-OV3 ovarian cancer cell line.[46] Although there are no definite conclusions to be drawn from

the increased (Table 3.2; fold change: +0.22) and decreased (Table 3.2; Figure 3.4: fold change: -0.33) levels of lysophosphatidylethanolamines detected in ET DKO mice, these changes could be related to the initiation of migration and invasion of HGSC to neighboring tissue during proliferation. Moreover, LysoPE(18:2) has recently been shown to aid in the detection of prostate cancer, another hormone-related cancer.[34]

3.4.4.3 Sphingoid Bases

Sphingoid bases, particularly sphingosine, have been long implicated in playing a role in cancer biology. One of the features in the discriminating subset was an unidentified $C_{18}H_{33}NO_2$ sphingoid base (Table 3.2; Figure 3.7). Sphingosine facilitates and triggers apoptosis, while sphingosine 1-phosphate promotes cellular survival/profileration in response to apoptotic stress.[71] Illuzzi *et al.* found that inhibiting sphingosine kinase, thereby suppressing the production of sphingosine 1-phosphate from sphingosine, reduced cell proliferation in ovarian cancer cell lines.[72] Research by Hong *et al.* suggested that sphingosine 1-phosphate is a vital component of cellular growth and adhesion modulation for ovarian cancer cell lines HEY and OCC1.[73] Increased levels of the $C_{18}H_{33}NO_2$ sphingoid base in ET DKO mice (Table 3.2; Figure 3.4; fold change: +1.23) therefore hint at a "biochemical fight response" toward the developing carcinoma. Supporting this hypothesis, Fan *et al.* detected increased plasma levels of the sphingolipids phytosphingosine, ganglioside, and ceramides in a cohort combining early- and late-stage EOC patients that were compared to healthy controls.[21] Additionally, Chen *et al.* observed an increase of phytosphingosine in the urine of a combined early- and late-stage EOC patient cohort.[18]

3.4.4.4 Bile Acids, Alcohols, & Derivatives

Bile acids cause apoptosis as well as oxidative stress in cells, resulting in the generation of reactive oxygen species (ROS) that subsequently damage DNA. Recurring DNA damage can increase mutation rates of onco- and tumor suppressor genes. Therefore, cells exposed to bile acids can proliferate, leading to carcinogenesis.[74] Two features in our discriminant feature subset were tentatively identified as bile acids and related substances (Table 3.2; Figure 3.7): 24-Nor-5β-cholane-triol and bilirubin. Serum levels of 24-Nor-5β-cholane-triol were found to be increased (Table 3.2; Figure 3.4; fold change: +0.46) in ET DKO mice. It is plausible that ROS generation by 24-Nor-5β-cholane-triol aided in the development of ovarian carcinoma in ET DKO mice. In agreement with these findings, Chen *et al.* identified 27-nor-5β-cholestane-3,7,12,24,25 pentol glucuronide as a serum biomarker for combined early- and late-stage EOC, which was also accompanied by increased levels of glycocholic acid.[25] Also supporting this finding is previous work by our group showing that bile acids, specifically isomers of 5β-chol-9(11)-en-24-oic acid, can aid in the detection of human HGSC.[22]

Although bilirubin is not a bile acid, it is a component of bile and the end product of heme catabolism. Furthermore, increased bilirubin has been associated with low cancer mortality.[75] Bilirubin acts as an endogenous antioxidant that reduces oxidative stress damage by scavenging peroxyl[76] and hydroxyl radicals.[77] Moreover, mutations in the tumor suppressor gene *p53* can be caused by ROS,[78] and *p53* mutations are found in many cancers, including ovarian cancer.[79, 80] Lowered concentration levels of bilirubin, as detected in the ET DKO mice (Table 3.2; Figure 3.4; fold change: -0.74), possibly imply diminished protection against ROS, thus leading to carcinogenesis. Supporting this

hypothesis, Chen *et al.* tentatively detected decreased levels of bilirubin in the serum of both early- and late-stage EOC patients, and increased levels of bilirubin mono- and di-glucuronic acids.[25]

3.4.4.5 Peptides

One member of the discriminant feature subset was tentatively identified as a tripeptide isomer of Ala-Trp-Ala or Lys-Cys-Pro (Table 3.2). Tandem MS data could not be acquired to confirm the tripeptide sequence as the precursor ion abundances were not sufficiently high. Blood serum levels of these peptides were decreased in ET DKO mice (Table 3.2; fold change: -0.52). Previous work by our group has also found that tripeptides, specifically Gln-His-Ala or its isomers, aided in human HGSC detection.[22] However, a survey of the literature revealed no mechanism by which small circulating peptides could play a biological role in ovarian cancer. Some proteases have been reported to be overexpressed in ovarian carcinoma,[81-83] but we would expect liberated peptides to have an increased concentration in ET DKO mice if their presence resulted from increased protease activity.

3.4.4.6 Phytochemicals – Terpenes

Three features in our discriminant subset were tentatively identified as terpenes (Table 3.2; Figure 3.7): a $C_{15}H_{24}O$ terpene derivative, uroterpenol-O-glucuronide, and an unidentified terpene ($C_{22}H_{32}O_2$). Many terpenes enter metabolic networks from food sources. Although both ET DKO and control mice were fed the same diet, it is plausible that the presence of HGSC in ET DKO mice altered the metabolism of dietary metabolites. Uroterpenol is a metabolic product of limonene, a terpene component of

citrus fruits, vegetables, herbs, and spices, and has been shown to lengthen tumor latency and decrease tumor yield in mice with mammary cancer.[84] Interestingly, uroterpenol-O-glucuronide has been found to have urinary excretion patterns resembling those of estriol, pregnanediol, and pregnanetriol, and increased excretions have been associated with increased placental, ovarian, or adrenal activity.[50] The decreased serum levels of uroterpenol-O-glucuronide observed in ET DKO mice (Table 3.2; fold change: -0.52) might thus relate to increased ovarian activity, or to decreased serum uroterpenol concentrations.

3.4.4.7 Phytochemicals – Lignans

One feature in our discriminant subset was an unidentified $C_{42}H_{58}O_4$ lignan. Lignans are phytoestrogens that are abundant in diets of people living in regions of low cancer incidence.[45] An increased dietary lignan intake, measured by urinary excretion, has been correlated with reduced breast cancer risk.[85] Ovarian cancer tumors contain type II estrogen binding sites,[86] and it has been suggested that lignans inhibit cancer cell growth by competing with estradiol for type II estrogen binding sites.[45] As there was a detected increase of the unidentified $C_{42}H_{58}O_4$ lignan (Table 3.2; fold change: +0.26) in the serum of ET DKO mice, it is possible that binding of lignans to type II estrogen sites was somehow decreased, thereby, not aiding in the inhibition of HGSC proliferation.

3.5 Ultra Performance Liquid Chromatography-Mass Spectrometry Characterization of Serum Metabolic Phenotypes of an Early-Stage Ovarian Cancer Pilot Patient Cohort

3.5.1 Multivariate Classification Performance

A total of 371 metabolic features (R_t, *m*/z pairs) were extracted from the UPLC-MS acquired metabolic profiles. A SVM model was built from these extracted features to discriminate between serum samples from early-stage OC patients and healthy women. The RFE method used in the model determined that a panel of 27 features was optimal to discriminate both sample classes with 100% accuracy, sensitivity, and specificity (Figure 3.8a). Of the 27 selected features, 2 were poorly resolved in the mass spectra and were thus removed from further consideration. The RFE method re-applied on this 25-feature panel led to a subgroup of 22 features that best differentiated early-stage OC patients from healthy women with 100% accuracy, sensitivity, and specificity (Figure 3.8b). Figure 3.9 shows the corresponding cross-validated "OC detection scores" obtained for both sample classes (H or D), which are separated by the optimal hyperplane (illustrated with a dotted line in the figure) that maximizes the margin between the most similar samples in each group.

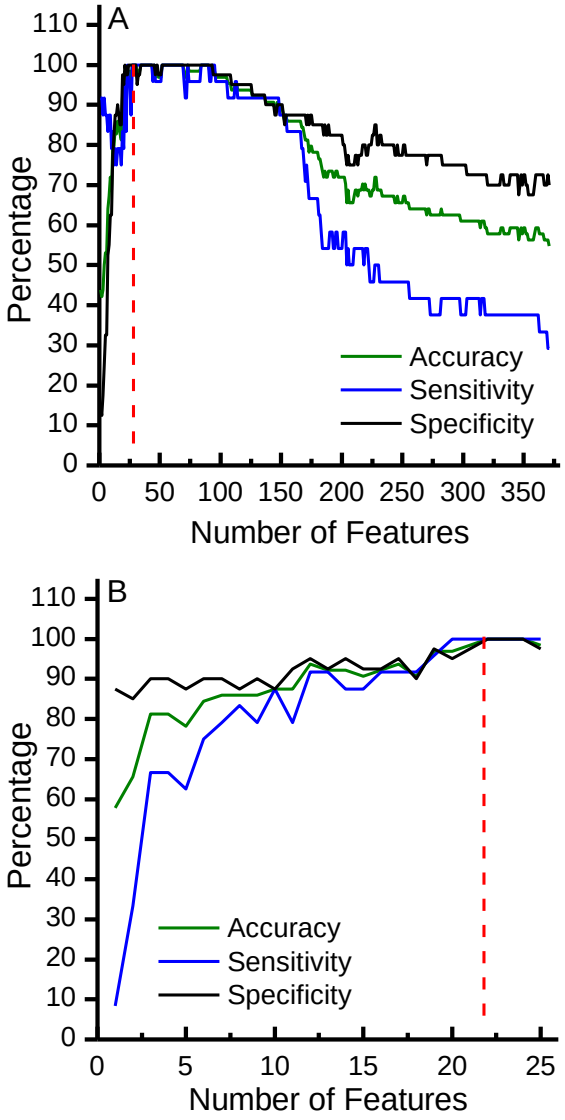

Figure 3.8: Evolution of classification accuracy, sensitivity, and specificity resulting from the recursive feature elimination (RFE) method using (A) all features and (B) the initial curated 25 discriminant feature set. The minimum discriminant feature set that maximizes the model performance characteristic is highlighted with a red dashed line. Accuracy values are represented by the green traces, sensitivity values are represented by the blue traces, and specificity values are represented by the black traces.

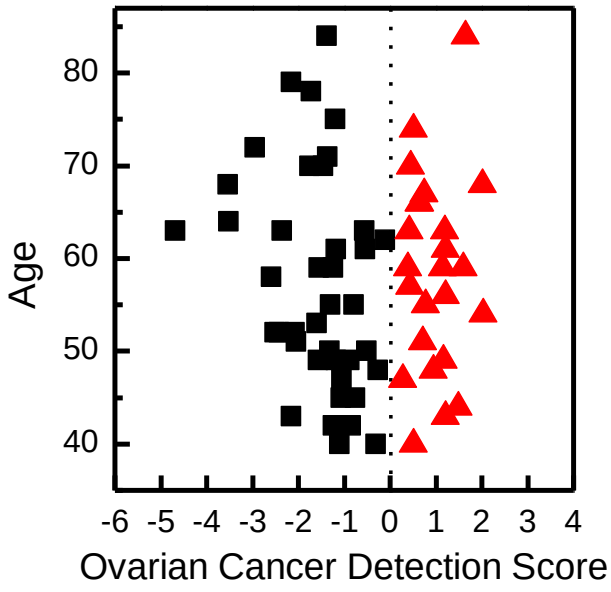

Figure 3.9: Visualization of ovarian cancer (OC) detection scores obtained by SVMs using the 22-feature discriminant panel. Black squares correspond to the healthy women and red triangles correspond to the early-stage OC patients. The dotted line shows the projection of the separating hyperplane: $\boldsymbol{wx}' + b = 0$.

The performance of this 22-feature discriminant panel was also investigated and compared to the initial set of 371 metabolic features by oPLS-DA. Misclassification plots were generated for both feature sets (Figure 3.10). Using the total number of features extracted from the UPLC-MS metabolic profiles, 4 latent variables interpreting 43.28% of the variance from the X- (feature peak areas) block and 73.59% of the variance from the Y- (class membership) block resulted in 32 sample misclassifications (Figure 3.10a). The cross-validated accuracy, sensitivity, and specificity were 49%, 64%, and 34%, respectively. However, the number of sample misclassifications was reduced to 4 with a 22-feature 4-latent variable oPLS-DA model, which interpreted 48.49% and 80.70% X-

and Y-block variance, respectively (Figure 3.10b). For this model, the cross-validated accuracy, sensitivity, and specificity were 94%, 97%, and 91%, respectively. These model performance characteristics increased by 33% or more upon variable selection, providing further evidence of the degree of useful information contained in the 22-feature discriminant panel.

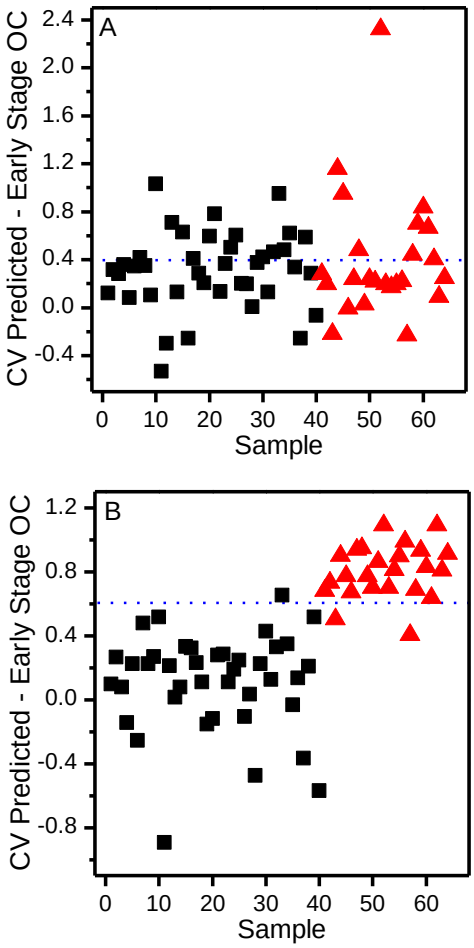

Figure 3.10: Orthogonal projection to latent structures-discriminant analysis (oPLS-DA) cross-validated (CV) prediction plots of early-stage ovarian cancer (OC) patients (red triangles) and healthy women (black squares). (A): OC CV prediction plot for the oPLS-DA model using the initial set of 371 spectral features. The cross-validated accuracy, sensitivity, and specificity of the model were 49%, 64%, and 34%, respectively. There were 32 misclassified samples. (B) OC CV prediction plot for the oPLS-DA model using the 22 discriminant feature panel. The cross-validated accuracy, sensitivity, and specificity of the model were 94%, 97%, and 91%, respectively. There were 4 misclassified samples.

3.5.2 Discriminant Metabolite Identification and Biological Relevance

Of the 22 RFE-selected discriminant features, the detected concentration levels of 9 were increased, whereas 13 were decreased in sera of early-stage OC patients (Figure 3.11). While only 5 of these concentration changes were univariately significant (Mann-Whitney U test, n = 64, $p \leq 0.05$; Figure 3.11), the co-varying concentrations of all 22 discriminant metabolic features were needed to distinguish the detected metabolomes of OC patients and healthy women. Moreover, the concentration levels of those features that were not univariately statistically different between early-stage OC patients and healthy women could feasibly be a result of the relatively small sample size used in this pilot study. Although metabolite identification was attempted for all 22 discriminant features once oPLS-DA confirmed the robustness of the SVM classification model, only 18 of the 22 metabolic features were identified by MS and MS/MS (Table 3.4). The majority of discriminant features were lipids and fatty acids (Table 3.4). Two of the features in Table 3.4 were tentatively matched to metabolites yet to be reported in mammals, such as xeniasterol-a, and their identity should therefore be considered only provisional. These tentative metabolite identifications are detailed in Table 3.5, and the metabolite chemical class is still listed in Table 3.4 (italicized).

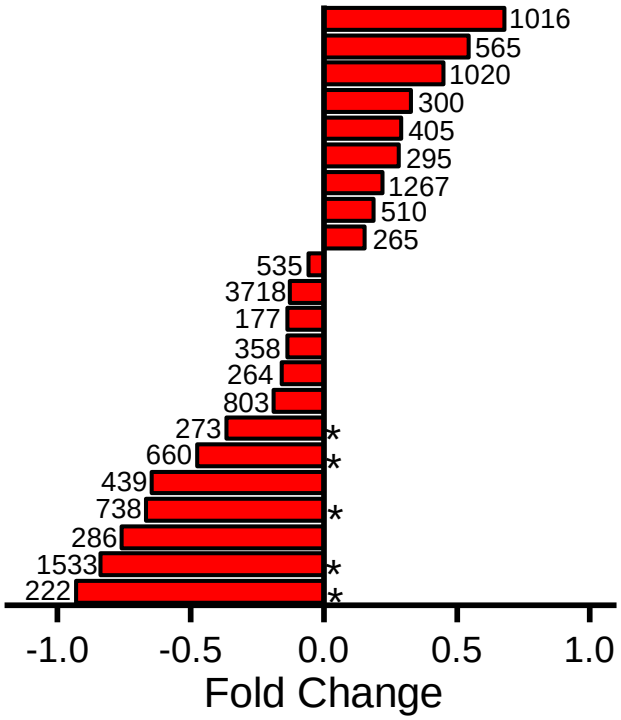

Figure 3.11: Fold change of average peak areas of each discriminant feature. Fold changes are calculated as the base 2 logarithm of the average peak area ratios for early-stage ovarian cancer samples and healthy women samples. Features are labeled with their codes. Features with statistically significant ($p \leq 0.05$) fold changes (calculated using Mann-Whitney U test) are additionally labeled with asterisks.

Table 3.4: Confirmed and Tentative Metabolites Identified as Discriminatory Between Early-Stage Ovarian Cancer Patients and Healthy Individuals. Metabolites confirmed by Tandem MS experiments are highlighted in bold font.

Feat. Code	Avg. Ret. time (min)[1]	Avg. m/*z*[1]	Fold Change[2]	P value[3]	Ion type	Tentative Metabolite Identification	Elemental Formula	Theoretical m/*z*	Δm (mDa)
295	**18.77**	**698.5570**	**0.25**	**0.10**	**[M-H]$^-$**	**Galactosylceramide (d18:1/16:0) Glucosylceramide (d18:1/16:0)**	**$C_{40}H_{77}NO_8$**	**698.5576**	**0.6**
738	11.41	479.2823	-0.64	0.014	[M-H]$^-$	PA(10:0/10:0)	$C_{23}H_{45}O_8P$	479.2779	4.4
1533	21.07	842.5910	-0.88	0.049	[M-H]$^-$	PE(22:2/22:6) PE(22:4/22:4)	$C_{49}H_{82}NO_8P$	842.5705	20.3
1016	13.26	355.2267	0.65	0.27	[M-H]$^-$	14-Hydroxy docosahexaenoic acid (14-HDoHE) 1-Formylneogrifolin	$C_{23}H_{32}O_3$	355.2279	1.2
						Pro Lys Leu[4] Lys Ile Pro[4]	$C_{17}H_{32}N_4O_4$	355.2351	8.4
286	20.56	764.5230	-0.79	0.095	[M-H]$^-$	PE(38:5)	$C_{43}H_{76}NO_8P$	764.5236	0.6
660	13.78	465.3571	-0.49	0.046	[M-H]$^-$	SM(d18:0/0:0) [LysoSM(d18:0)]	$C_{23}H_{51}N_2O_5P$	465.3463	10.8
358	15.81	491.3730	-0.15	0.46	[M-H]$^-$	*Polyhydroxysterol*	$C_{30}H_{52}O_5$	491.3742	1.2
265	**13.13**	**599.3195**	**0.18**	**0.93**	**[M-H]$^-$**	**PI(18:0/0:0) [LysoPI(18:0)]**	**$C_{27}H_{53}O_{12}P$**	**599.3202**	**0.7**
222	**21.10**	**722.5123**	**-0.96**	**0.016**	**[M-H]$^-$**	**PE(P-16:0/20:4)**	**$C_{41}H_{74}NO_7P$**	**722.5130**	**0.7**
264	13.92	772.5492	-0.17	0.087	[M-H]$^-$	PE(P-18:1/22:6)	$C_{45}H_{76}NO_7P$	772.5287	20.5
1267	**16.47**	**833.5183**	**0.19**	**0.57**	**[M-H]$^-$**	**PI(16:0/18:2) PI(16:1/18:1)**	**$C_{43}H_{79}O_{13}P$**	**833.5186**	**0.3**

Table 3.4 (continued).

177	**14.99**	**305.2478**	**-0.11**	**0.41**	**[M-H]$^-$**	**5,8,11-Eicosatrienoic acid (Mead acid) 8,11,14-Eicosatrienoic acid (Dihomo-γ-Linolenic Acid)**	**$C_{20}H_{34}O_2$**	**305.2486**	**0.8**
1020	12.96	557.3681	0.44	0.12	[M-H]$^-$	Gln Arg Lys Lys[4]	$C_{23}H_{46}N_{10}O_6$	557.3529	15.2
405	18.31	602.4552	0.29	0.17	[M-H]$^-$	N/A	-	-	-
300	**18.51**	**860.6101**	**0.31**	**0.089**	**[M-H]$^-$**	**Lactosylceramide (d18:1/16:0) Galabiosylceramide (d18:1/16:0)**	**$C_{46}H_{87}NO_{13}$**	**860.6105**	**0.4**
3718	0.76	609.2308	-0.17	0.86	[M-2H]$^{2-}$	N/A	-	-	-
273	14.19	774.5648	-0.31	0.025	[M-H]$^-$	PE(P-18:1/22:5) PE(P-18:0/22:6)	$C_{45}H_{78}NO_7P$	774.5443	20.5
439	18.11	379.3571	-0.64	0.082	[M-H]$^-$	*Monounsaturated fatty acid*	$C_{25}H_{48}O_2$	379.3581	1.0
510	16.10	1151.7067	0.16	0.32	[M-H]$^-$	Ganglioside GM3 (d18:1/16:0)	$C_{57}H_{104}N_2O_{21}$	1151.7059	0.8
803	19.27	797.6533	-0.15	0.95	[M-H]$^-$	SM(d18:2/23:0) SM(d17:1/24:1)	$C_{46}H_{91}N_2O_6P$	797.6542	0.9
535	13.91	577.4225	-0.05	0.63	[M-H]$^-$	2-Hexaprenyl-3-methyl-5-hydroxy-6-methoxy-1,4-benzoquinol	$C_{38}H_{58}O_4$	577.4262	3.7
565	**16.24**	**807.5037**	**0.54**	**0.89**	**[M-H]$^-$**	**PI(16:0/16:1)**	**$C_{41}H_{77}O_{13}P$**	**807.5029**	**0.8**

[1]These reported values were obtained after data processing with MZmine 2.0 software which generates average values calculated using all samples.
[2]Fold change was calculated as the base 2 logarithm of the average peak area ratios for OC samples and control samples.
[3]P values calculated using Mann-Whitney U test.
[4]Other polypeptide isomers are not listed; tandem MS data could not be acquired to confirm true isomer identity as their ion abundances were not high enough for quadrupole selection.

Table 3.5: Metabolites Not Reported in Mammals Identified as Discriminatory Between Early-Stage OC Patients and Healthy Individuals.

Feat. Code	Avg. Ret. time (min)[1]	Avg. m/z[1]	Fold Change[2]	P value[3]	Ion type	Tentative Metabolite Identification	Elemental Formula	Theoretical m/z	Δm (mDa)
						(Polyhydroxysterol)			
358	15.81	491.3730	-0.15	0.46	$[M-H]^-$	Xeniasterol-a 25-acetoxy-ergosta-3β,5α,6β-triol	$C_{30}H_{52}O_5$	491.3742	1.2
						(Monounsaturated fatty acid)			
439	18.11	379.3571	-0.64	0.082	$[M-H]^-$	Pentacosenoic acid Mycolipenic acid	$C_{25}H_{48}O_2$	379.3581	1.0

[1]These reported values were obtained after data processing with MZmine 2.0 software which generates average values calculated using all samples.
[2]Fold change was calculated as the base 2 logarithm of the average peak area ratios for OC samples and control samples.
[3]P values calculated using Mann-Whitney U test.

Five discriminant metabolites are sphingolipids involved in ceramide-mediated metabolic pathways. Sphingolipid metabolism alterations have been implicated in cancer pathogenesis and progression.[21, 87] Ceramides aid in cell cycle arrest, terminal cell differentiation, and apoptosis,[88, 89] but have been reported to be decreased in OC tumors compared to normal ovarian tissues,[90] thereby, aiding OC proliferation. Decreases of two sphingomyelins (SM(d18:0/0:0) and SM(d18:2/23:0) or SM(d17:1/24:1)) and increases of a cerebroside (glucosylceramide (d18:1/16:0) or galactosylceramide (d18:1/16:0)), Ganglioside GM3, and a globoside (lactosylceramide (d18:1/16:0) or galabiosylceramide (d18:1/16:0)) were observed in the sera of early-stage OC patients (Table 3.4; Figure 3.11). Sphingomyelins can be hydrolyzed to form ceramides, which may account for their observed decrease. The resultant ceramides can be incorporated into cerebrosides (precursors to gangliosides), thereby supporting their observed increase in the early-stage OC patients. Supporting our observations, elevated levels of glucocerebrosides in the multi-drug resistant OVCAR-3, an OC cell line, have been reported.[91] Additionally, Fan et al. found that the plasma concentration levels of a ganglioside were increased in combined early- and late-stage OC patients when compared to healthy women.[21] Our findings suggest that the early-stage OC patients included in this study have decreased levels of ceramides, which are essential in the apoptotic response of cancer cells to death inducers such as hypoxia and DNA damage.[92] Sphingolipid metabolism was also found to be modified in *Dicer* $^{flox/flox}$ *Pten* $^{flox/flox}$ *Amhr2* $^{cre/+}$ double-knockout mice with early-stage ovarian cancer.[93]

Phosphatidylinositols (Table 3.4; Figure 3.11) were identified as discriminant metabolites with increased levels in the early-stage OC patients, in agreement with the

fact that the phosphatidylinositol 3-kinase (PI3K)-AKT pathway has long been implicated in ovarian carcinogensis.[94] The activation of the PI3K oncogene, whose encoded protein phosphorylates phosphatidylinositols, has been linked to increased cell survival due to the suppression of apoptosis.[69, 70] Although our data only represents a small snapshot of the underlying biochemical processes occurring after the onset of OC, it is plausible that activity of *PTEN*, a PI3K-AKT pathway regulator, was enhanced in an attempt to stifle cellular growth of OC cells, thereby, increasing the levels of phosphatidylinositols in the early-stage OC patients. While *PTEN* levels are known to decline during the progression of OC,[95] they have a strong inverse correlation with PI3K-AKT activity.[96] A second group of phospholipids with decreased levels in the early-stage OC patients was identified in the discriminant panel as phosphatidylethanolamines (Table 3.4; Figure 3.11). Previous work by our group found that a collection of phosphatidylethanolamines aided in combined early- and late-stage OC detection.[22] Additionally, alterations in the metabolism of phosphatidylethanolamines were also detected in *Dicer* $^{flox/flox}$ *Pten* $^{flox/flox}$ *Amhr2* $^{cre/+}$ double-knockout mice with early-stage ovarian cancer.[93] Human phosphatidylethanolamine-binding protein 4 (hPEBP4), a known protein overly expressed in the CaoV-3 OC cell line,[97, 98] has been shown to inhibit tumor necrosis factor-α (TNF-α) induced apoptosis.[99] Binding of hPEBP4 to phosphatidylethanolamines could account for their decreased levels in the early-stage OC patients. Phosphatidic acid was an additional identified discriminant metabolite (Table 3.4; Figure 3.11). This compound plays a direct role in the metabolism of lysophosphatidic acid (LPA),which activates proliferative growth signaling cascades and suppresses apoptotic cell death in OC when it is altered.[100] Since these phospholipids play

a role in maintaining cellular membrane integrity, alterations in their metabolism also suggest membrane degradation or morphological changes in OC-affected cells. Furthermore, membrane phospholipid translocation is an early indicator of apoptosis.[101, 102] These findings add to the growing body of scientific evidence that shows apoptosis is gravely disturbed during ovarian carcinogenesis.

Animal and *in vitro* studies have shown that n-3 fatty acids, such as docosahexaenoic acid (DHA), inhibit carcinogenesis in hormone-related cancers.[103] Though several mechanisms by which this inhibition occurs have been postulated, one is *via* suppression of arachidonic acid (AA)-derived eicosanoids.[104] These eicosanoids can alter immune responses to cancer cells in addition to inhibiting apoptosis,[105] stimulating cellular proliferation and metastasis,[106] and promoting tumor angiogenesis.[107] 14-Hydroxy docosahexaenoic acid (14-HDoHE), an oxidation product of DHA metabolism, was detected in higher concentrations in the early-stage OC patients (Table 3.4; Figure 3.11), thus, suggesting that DHA may not be able to inhibit carcinogenesis by suppression of AA-derived eicosanoids during the early stages of OC. Furthermore, dihomo-γ-linolenic acid (DGLA), which is involved in the biosynthesis of AA, was detected at lower levels in the early-stage OC patients (Table 3.4; Figure 3.11), hinting that AA-derived eicosanoids may be increased in the early stages of ovarian carcinogenesis if the decrease of DGLA is due to increased synthesis of AA. DGLA is also an activator of peroxisome proliferator-activated receptor-δ (PPARδ), which is a cellular proliferation inducer and promoter of cancer progression.[104, 108]

2-Hexaprenyl-3-methyl-5-hydroxy-6-methoxy-1,4-benzoquinol, a ubiquinone derivative, had lower levels in the early-stage OC patients (Table 3.4; Figure 3.11).

Ubiquinone is an electron carrier and, thus, an antioxidant. Studies have shown that ubiquinone levels are increased in precancerous tissue changes and decrease below normal tissue levels when cancer is fully developed. These changes imply that levels of ubiquinone are affected as cells respond to the increased oxidative stress caused by carcinogenesis.[109, 110]

3.5.3 Discriminant Metabolite Differences Between Early-Stage Papillary Serous and Endometrioid Ovarian Carcinoma

The early-stage OC patient cohort in this study was comprised of patients having either papillary serous (n = 12) or endometrioid (n = 12) histological subtypes of OC (Table 3.6).

Table 3.6. Summary of Patient Cohort Characteristics.

Ovarian Histopathology	No. of Patients	No. of Stage I	No. of Stage II	Age Range (at Surgery)	Average Age (at Surgery)
Papillary serous carcinoma	12	7	5	49-84	63
Endometrioid carcinoma	12	11	1	40-70	54
Within normal limits	40	N/A	N/A	40-84	57

Serous ovarian carcinoma is the most common type of epithelial OC, while endometrioid accounts for 10-20% of all epithelial OC cases.[111] Serous OC mainly originates from the mesothelial surface lining of the ovarian cortex or cortical inclusion cysts, and the onset

of endometrioid OC is greatly associated with ovarian endometriosis.[112, 113] Since microarray gene expression analysis has shown that the etiology of OC histological subtypes may be different due to the activation and deregulation of different pathways,[113] metabolome differences were explored in these subtypes as well within the 22-feature panel. However, only 2 discriminant features of the panel presented significantly different concentration levels between serous and endometrioid OC. Because the SVM model aimed to differentiate both OC histological subtype groups from the healthy women, they were grouped together to find common features for sample discrimination, and presumably, the features selected by RFE would not differ between the two OC histological subtypes. Feature 1016 (14-HDoHE) had higher concentration levels in serous OC patients (fold change: 2.5), whereas feature 3718 (unidentified) had lower concentration levels in serous OC patients (fold change: -1.1) (Figure 3.12). Perhaps DHA is metabolized at a higher rate in early-stage papillary serous OC patients, thereby, limiting its ability to incorporate into cellular membranes and replace AA— the mechanism by which DHA suppresses the biosynthesis of AA-derived eicosanoids.[104]

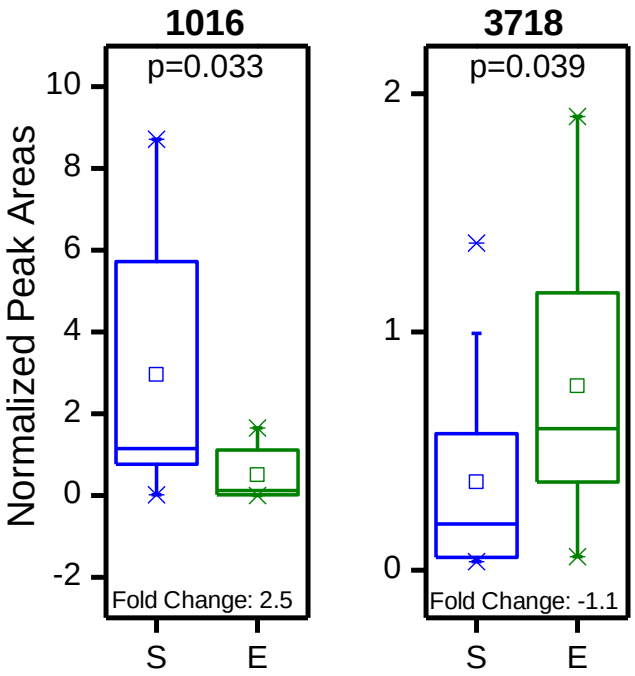

Figure 3.12: Discriminatory spectral features having statistically significant univariate changes between papillary serous ovarian cancer (S) patients (n = 12) and endometrioid ovarian cancer (E) patients (n = 12). *P* values were calculated using the Mann−Whitney U test. Box plots with mean (square), median, upper and lower quartile, outliers, and minimum and maximum (whiskers) data values are displayed. Feature ID numbers are indicated on top of each case. Fold change was calculated as the base 2 logarithm of the average peak area ratios for early-stage papillary serous ovarian cancer samples and early-stage endometrioid ovarian cancer samples.

3.6 Conclusion

This chapter details the success of two systems used to investigate metabolic patterns for early detection of OC. In the first system, early-stage HGSC was successfully detected in a *Dicer-Pten* DKO mice model utilizing UPLC-MS untargeted metabolic fingerprinting. After down-selection of spectral features with maximum discriminatory power, eighteen metabolites differentiated DKO mice with early-stage tumors from

control mice with 100% accuracy, sensitivity and specificity. Altered metabolic pathways included those of fatty acids, bile acids and alcohols, glycerophospholipids, peptides, and phytochemicals. These alterations impact cellular energy storage and membrane stability, as well as defenses against oxidative stress. This work is the first step toward understanding the underlying metabolic changes resulting from the progression of early HGSCs originating in the fallopian tube, but an exploration of the overlap between DKO mice and humans at early stages of HGSC development is needed. Building upon previous work, the second system studied showed that early-stage OC detection in humans employing UPLC-MS untargeted metabolic profiling is feasible. After down-selection of spectral features with maximum discriminatory power by RFE, 22 metabolites differentiated 24 early-stage OC patients from 40 healthy women with 100% cross-validated accuracy, sensitivity, and specificity. Although this is a small pilot study—poor early diagnosis currently complicates collection of large patient cohorts for more detailed studies—it demonstrates that metabolic information can be useful for detecting early-stage OC. Larger, more focused studies will be needed to further validate our findings.

3.7 References

(1) Siegel, R.; Naishadham, D.; Jemal, A., Cancer statistics, 2013. *CA Cancer J. Clin.* **2013,** *63* (1), 11-30.

(2) Cho, K. R.; Shih, I., Ovarian cancer. *Annu. Rev. Pathol-Mech.* **2009,** *4*, 287-313.

(3) Bast, R. C., Jr.; Hennessy, B.; Mills, G. B., The biology of ovarian cancer: new opportunities for translation. *Nat. Rev. Cancer* **2009,** *9* (6), 415-428.

(4) TCGA, Integrated genomic analyses of ovarian carcinoma. *Nature* **2011,** *474* (7353), 609-615.

(5) Bhoola, S.; Hoskins, W. J., Diagnosis and management of epithelial ovarian cancer. *Obstet. Gynecol.* **2006,** *107* (6), 1399-1410.

(6) Seidman, J. D.; Zhao, P.; Yemelyanova, A., "Primary peritoneal" high-grade serous carcinoma is very likely metastatic from serous tubal intraepithelial carcinoma: assessing the new paradigm of ovarian and pelvic serous carcinogenesis and its implications for screening for ovarian cancer. *Gynecol. Oncol.* **2011,** *120* (3), 470-473.

(7) Siegel, R.; Ma, J.; Zou, Z.; Jemal, A., Cancer statistics, 2014. *CA Cancer J. Clin.* **2014,** *64* (1), 9-29.

(8) Williams, T. I.; Toups, K. L.; Saggese, D. A.; Kalli, K. R.; Cliby, W. A.; Muddiman, D. C., Epithelial ovarian cancer: Disease etiology, treatment, detection, and investigational gene, metabolite, and protein biomarkers. *J. Proteome Res.* **2007,** *6* (8), 2936-2962.

(9) Bristow, R. E.; Hodeib, M.; Smith, A.; Chan, D. W.; Zhang, Z.; Fung, E. T.; Tewari, K. S.; Munroe, D. G.; Ueland, F. R., Impact of a multivariate index assay on referral patterns for surgical management of an adnexal mass. *Am. J. Obstet. Gynecol.* **2013,** *209* (6), 581.e1-581.e8.

(10) Ueland, F. R.; Desimone, C. P.; Seamon, L. G.; Miller, R. A.; Goodrich, S.; Podzielinski, I.; Sokoll, L.; Smith, A.; van Nagell, J. R.; Zhang, Z., Effectiveness of a multivariate index assay in the preoperative assessment of ovarian tumors. *Obstet. Gynecol.* **2011,** *117* (6), 1289-1297.

(11) Zhang, Z., An in vitro diagnostic multivariate index assay (IVDMIA) for ovarian cancer: harvesting the power of multiple biomarkers. *Rev. Obstet. Gynecol.* **2012,** *5* (1), 35-41.

(12) Gynecologists, A. C. o. O. a., Committee Opinion 477: The role of the obstetrician-gynecologist in the early detection of epithelial ovarian cancer. *Obstet. Gynecol.* **2011,** *117* (3), 742-746.

(13) Piek, J. M.; van Diest, P. J.; Zweemer, R. P.; Jansen, J. W.; Poort-Keesom, R. J.; Menko, F. H.; Gille, J. J.; Jongsma, A. P.; Pals, G.; Kenemans, P.; Verheijen, R. H., Dysplastic changes in prophylactically removed Fallopian tubes of women predisposed to developing ovarian cancer. *J. Pathol.* **2001,** *195* (4), 451-456.

(14) Crum, C. P.; Drapkin, R.; Kindelberger, D.; Medeiros, F.; Miron, A.; Lee, Y., Lessons from BRCA: The tubal fimbria emerges as an origin for pelvic serous cancer. *Clin. Med. Res.* **2007,** *5* (1), 35-44.

(15) Kurman, R. J.; Shih, I., The origin and pathogenesis of epithelial ovarian cancer: a proposed unifying theory. *Am. J. Surg. Pathol.* **2010,** *34* (3), 433-443.

(16) Kim, J.; Coffey, D. M.; Creighton, C. J.; Yu, Z.; Hawkins, S. M.; Matzuk, M. M., High-grade serous ovarian cancer arises from fallopian tube in a mouse model. *Proc. Natl. Acad. Sci. USA* **2012,** *109* (10), 3921-3926.

(17) Zhang, T.; Wu, X. Y.; Ke, C. F.; Yin, M. Z.; Li, Z. Z.; Fan, L. J.; Zhang, W.; Zhang, H. Y.; Zhao, F. L.; Zhou, X. H.; Lou, G.; Li, K., Identification of potential biomarkers for ovarian cancer by urinary metabolomic profiling. *J. Proteome Res.* **2013,** *12* (1), 505-512.

(18) Chen, J.; Zhou, L. N.; Zhang, X. Y.; Lu, X.; Cao, R.; Xu, C. J.; Xu, G. W., Urinary hydrophilic and hydrophobic metabolic profiling based on liquid chromatography-mass spectrometry methods: differential metabolite discovery specific to ovarian cancer. *Electrophoresis* **2012,** *33* (22), 3361-3369.

(19) Slupsky, C. M.; Steed, H.; Wells, T. H.; Dabbs, K.; Schepansky, A.; Capstick, V.; Faught, W.; Sawyer, M. B., Urine metabolite analysis offers potential early diagnosis of ovarian and breast cancers. *Clin. Cancer. Res.* **2010,** *16* (23), 5835-5841.

(20) Woo, H. M.; Kim, K. M.; Choi, M. H.; Jung, B. H.; Lee, J.; Kong, G.; Nam, S. J.; Kim, S.; Bai, S. W.; Chung, B. C., Mass spectrometry based metabolomic approaches in urinary biomarker study of women's cancers. *Clin. Chim. Acta* **2009,** *400* (1), 63-69.

(21) Fan, L. J.; Zhang, W.; Yin, M. Z.; Zhang, T.; Wu, X. Y.; Zhang, H. Y.; Sun, M.; Li, Z. Z.; Hou, Y.; Zhou, X. H.; Lou, G.; Li, K., Identification of metabolic biomarkers to diagnose epithelial ovarian cancer using a UPLC/QTOF/MS platform. *Acta Oncol.* **2012,** *51* (4), 473-479.

(22) Guan, W.; Zhou, M.; Hampton, C.; Benigno, B.; Walker, L. D.; Gray, A.; McDonald, J.; Fernandez, F., Ovarian cancer detection from metabolomic liquid

chromatography/mass spectrometry data by support vector machines. *BMC Bioinformatics* **2009,** *10*, 259.

(23) Odunsi, K.; Wollman, R. M.; Ambrosone, C. B.; Hutson, A.; McCann, S. E.; Tammela, J.; Geisler, J. P.; Miller, G.; Sellers, T.; Cliby, W.; Qian, F.; Keitz, B.; Intengan, M.; Lele, S.; Alderfer, J. L., Detection of epithelial ovarian cancer using ^{1}H-NMR-based metabonomics. *Int. J. Cancer* **2005,** *113* (5), 782-788.

(24) Garcia, E.; Andrews, C.; Hua, J.; Kim, H. L.; Sukumaran, D. K.; Szyperski, T.; Odunsi, K., Diagnosis of early stage ovarian cancer by ^{1}H NMR metabonomics of serum explored by use of a microflow NMR probe. *J. Proteome Res.* **2011,** *10* (4), 1765-1771.

(25) Chen, J.; Zhang, X. Y.; Cao, R.; Lu, X.; Zhao, S. M.; Fekete, A.; Huang, Q.; Schmitt-Kopplin, P.; Wang, Y. S.; Xu, Z. L.; Wan, X. P.; Wu, X. H.; Zhao, N. Q.; Xu, C. J.; Xu, G. W., Serum 27-nor-5 beta-Cholestane-3,7,12,24,25 Pentol Glucuronide discovered by metabolomics as potential diagnostic biomarker for epithelium ovarian cancer. *J. Proteome Res.* **2011,** *10* (5), 2625-2632.

(26) Zhou, M. S.; Guan, W.; Walker, L. D.; Mezencev, R.; Benigno, B. B.; Gray, A.; Fernandez, F. M.; McDonald, J. F., Rapid mass spectrometric metabolic profiling of blood sera detects ovarian cancer with high accuracy. *Cancer Epidemiol. Biomarkers Prev.* **2010,** *19* (9), 2262-2271.

(27) Denkert, C.; Budczies, J.; Kind, T.; Weichert, W.; Tablack, P.; Sehouli, J.; Niesporek, S.; Konsgen, D.; Dietel, M.; Fiehn, O., Mass spectrometry-based metabolic profiling reveals different metabolite patterns in invasive ovarian carcinomas and ovarian borderline tumors. *Cancer Res.* **2006,** *66* (22), 10795-10804.

(28) Ben Sellem, D.; Elbayed, K.; Neuville, A.; Moussallieh, F.-M.; Lang-Averous, G.; Piotto, M.; Bellocq, J.-P.; Namer, I., Metabolomic characterization of ovarian epithelial carcinomas by HRMAS-NMR spectroscopy. *J. Oncol.* **2011,** *2011*, 8 pages.

(29) Fong, M. Y.; McDunn, J.; Kakar, S. S., Identification of metabolites in the normal ovary and their transformation in primary and metastatic ovarian cancer. *PLoS One* **2011,** *6* (5), e19963.

(30) Moyer, V. A., Screening for ovarian cancer: U.S. preventive services task force reaffirmation recommendation statement. *Ann. Intern. Med.* **2012,** *157* (12), 900-904.

(31) Pluskal, T.; Castillo, S.; Villar-Briones, A.; Orešič, M., MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* **2010,** *11*, 395.

(32) Brereton, R. G.; Lloyd, G. R., Support Vector Machines for classification and regression. *Analyst* **2010,** *135* (2), 230-267.

(33) Fan, R.-E.; Chang, K.-W.; Hsieh, C.-J.; Wang, X.-R.; Lin, C.-J., LIBLINEAR: A library for large linear classification. *J. Mach. Learn. Res.* **2008,** *9*, 1871-1874.

(34) Zang, X.; Jones, C. M.; Long, T. Q.; Monge, M. E.; Zhou, M.; Walker, L. D.; Mezencev, R.; Gray, A.; McDonald, J. F.; Fernandez, F. M., Feasibility of detecting prostate cancer by ultra performance liquid chromatography−mass spectrometry serum metabolomics. *J. Proteome Res.* **2014,** *13* (7), 3444-3454.

(35) Lokhov, P. G.; Kharybin, O. N.; Archakov, A. I., Diagnosis of lung cancer based on direct-infusion electrospray mass spectrometry of blood plasma metabolites. *Int. J. Mass spectrom.* **2012,** *309*, 200-205.

(36) Frickenschmidt, A.; Fröhlich, H.; Bullinger, D.; Zell, A.; Laufer, S.; Gleiter, C. H.; Liebich, H.; Kammerer, B., Metabonomics in cancer diagnosis: mass spectrometry-based profiling of urinary nucleosides from breast cancer patients. *Biomarkers* **2008,** *13* (4), 435-449.

(37) Zomer, S.; Brereton, R. G.; Carter, J. F.; Eckers, C., Support vector machines for the discrimination of analytical chemical data: application to the determination of tablet production by pyrolysis-gas chromatography-mass spectrometry. *Analyst* **2004,** *129* (2), 175-181.

(38) Smith, C. A.; O'Maille, G.; Want, E. J.; Qin, C.; Trauger, S. A.; Brandon, T. R.; Custodio, D. E.; Abagyan, R.; Siuzdak, G., METLIN: a metabolite mass spectral database. *Ther. Drug Monit.* **2005,** *27* (6), 747-751.

(39) Sud, M.; Fahy, E.; Cotter, D.; Brown, A.; Dennis, E. A.; Glass, C. K.; Merrill, A. H.; Murphy, R. C.; Raetz, C. R. H.; Russell, D. W.; Subramaniam, S., LMSD: LIPID MAPS structure database. *Nucleic Acids Res.* **2007,** *35* (suppl 1), D527-D532.

(40) Wishart, D. S.; Tzur, D.; Knox, C.; Eisner, R.; Guo, A. C.; Young, N.; Cheng, D.; Jewell, K.; Arndt, D.; Sawhney, S., HMDB: The human metabolome database. *Nucleic Acids Res.* **2007,** *35* (suppl 1), D521-D526.

(41) Zhou, B.; Wang, J.; Ressom, H. W., MetaboSearch: tool for mass-based metabolite identification using multiple databases. *PLoS One* **2012,** *7* (6), e40096.

(42) Douglas, D. J.; Konenkov, N. V., Ion source emittance influence on the transmission of a quadrupole operated in the second stability region. *J. Am. Soc. Mass. Spectrom.* **1998,** *9* (10), 1074-1080.

(43) Horai, H.; Arita, M.; Kanaya, S.; Nihei, Y.; Ikeda, T.; Suwa, K.; Ojima, Y.; Tanaka, K.; Tanaka, S.; Aoshima, K., MassBank: a public repository for sharing mass spectral data for life sciences. *J. Mass Spectrom.* **2010,** *45* (7), 703-714.

(44) Ulmer, H.; Borena, W.; Rapp, K.; Klenk, J.; Strasak, A.; Diem, G.; Concin, H.; Nagel, G., Serum triglyceride concentrations and cancer risk in a large cohort study in Austria. *Br. J. Cancer* **2009,** *101* (7), 1202-1206.

(45) Adlercreutz, H.; Mousavi, Y.; Clark, J.; Höckerstedt, K.; Hämäläinen, E.; Wähälä, K.; Mäkelä, T.; Hase, T., Dietary phytoestrogens and cancer: *in vitro* and *in vivo* studies. *J. Steroid Biochem. Mol. Biol.* **1992,** *41* (3), 331-337.

(46) Park, K. S.; Lee, H. Y.; Lee, S. Y.; Kim, M.-K.; Kim, S. D.; Kim, J. M.; Yun, J.; Im, D.-S.; Bae, Y.-S., Lysophosphatidylethanolamine stimulates chemotactic migration and cellular invasion in SK-OV3 human ovarian cancer cells: involvement of pertussis toxin-sensitive G-protein coupled receptor. *FEBS Lett.* **2007,** *581* (23), 4411-4416.

(47) Iorio, E.; Ricci, A.; Bagnoli, M.; Pisanu, M. E.; Castellano, G.; Di Vito, M.; Venturini, E.; Glunde, K.; Bhujwalla, Z. M.; Mezzanzanica, D., Activation of phosphatidylcholine cycle enzymes in human epithelial ovarian cancer cells. *Cancer Res.* **2010,** *70* (5), 2126-2135.

(48) Ye, B.; Aponte, M.; Dai, Y.; Li, L.; Ho, M.-C. D.; Vitonis, A.; Edwards, D.; Huang, T.-N.; Cramer, D. W., *Ginkgo biloba* and ovarian cancer prevention: epidemiological and biological evidence. *Cancer Lett.* **2007,** *251* (1), 43-52.

(49) Gercel-Taylor, C.; Doering, D. L.; Kraemer, F. B.; Taylor, D. D., Aberrations in normal systemic lipid metabolism in ovarian cancer patients. *Gynecol. Oncol.* **1996,** *60* (1), 35-41.

(50) Smith, O. W.; Wade, A.; Dean, F., Uroterpenol, a pettenkofer chromogen of dietary origin and a common constituent of human urine. *J. Endocrinol.* **1969,** *45* (1), 17-28.

(51) Menendez, J. A.; Papadimitropoulou, A.; Vellon, L.; Lupu, R., A genomic explanation connecting "Mediterranean diet", olive oil and cancer: oleic acid, the main monounsaturated fatty acid of olive oil, induces formation of inhibitory "PEA3 transcription factor-PEA3 DNA binding site" complexes at the Her-2/ *neu* (*erb*B-2) oncogene promoter in breast, ovarian and stomach cancer cells. *Eur. J. Cancer* **2006,** *42* (15), 2425-2432.

(52) Shayesteh, L.; Lu, Y.; Kuo, W.-L.; Baldocchi, R.; Godfrey, T.; Collins, C.; Pinkel, D.; Powell, B.; Mills, G. B.; Gray, J. W., PIK3CA is implicated as an oncogene in ovarian cancer. *Nat. Genet.* **1999,** *21* (1), 99-102.

(53) Kanehisa, M.; Goto, S., KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **2000,** *28* (1), 27-30.

(54) Caspi, R.; Altman, T.; Billington, R.; Dreher, K.; Foerster, H.; Fulcher, C. A.; Holland, T. A.; Keseler, I. M.; Kothari, A.; Kubo, A., The MetaCyc database of

metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res.* **2014,** *42* (D1), D459-D471.

(55) Risch, H. A.; Jain, M.; Marrett, L. D.; Howe, G. R., Dietary fat intake and risk of epithelial ovarian cancer. *J. Natl. Cancer Inst.* **1994,** *86* (18), 1409-1415.

(56) DeBerardinis, R. J.; Sayed, N.; Ditsworth, D.; Thompson, C. B., Brick by brick: metabolism and tumor cell growth. *Curr. Opin. Genet. Dev.* **2008,** *18* (1), 54-61.

(57) Argiles, J. M.; Alvarez, B.; López-Soriano, F. J., The metabolic basis of cancer cachexia. *Med. Res. Rev.* **1997,** *17* (5), 477-498.

(58) O'Donnell, A. J. M.; Macleod, K. G.; Burns, D. J.; Smyth, J. F.; Langdon, S. P., Estrogen receptor-α mediates gene expression changes and growth response in ovarian cancer cells exposed to estrogen. *Endocr. Relat. Cancer* **2005,** *12* (4), 851-866.

(59) Kim, H.-J.; Kalkhoff, R., Sex steroid influence on triglyceride metabolism. *J. Clin. Invest.* **1975,** *56* (4), 888.

(60) Park, J. R.; Cho, B. S., Changes in plasma lipids, lipoproteins, triglyceride secretion and removal in chicks with estrogen implants. *Lipids* **1988,** *23* (4), 327-333.

(61) Kudzma, D. J.; Hegstad, P. M.; Stoll, R. E., The chick as a laboratory model for the study of estrogen-induced hyperlipidemia. *Metabolism* **1973,** *22* (3), 423-434.

(62) Hazzard, W. R.; Spiger, M. J.; Bagdade, J. D.; Bierman, E. L., Studies on the mechanism of increased plasma triglyceride levels induced by oral contraceptives. *New Engl. J. Med.* **1969,** *280* (9), 471-474.

(63) O'Brien, T.; Nguyen, T. T. In *Lipids and lipoproteins in women*, Mayo Clin. Proc., 1997; Elsevier: 1997; pp 235-244.

(64) Ha, Y.; Grimm, N.; Pariza, M., Anticarcinogens from fried ground beef: heat-altered derivatives of linoleic acid. *Carcinogenesis* **1987,** *8* (12), 1881-1887.

(65) Ip, C.; Chin, S. F.; Scimeca, J. A.; Pariza, M. W., Mammary cancer prevention by conjugated dienoic derivative of linoleic acid. *Cancer Res.* **1991,** *51* (22), 6118-6124.

(66) Ando, A.; Ogawa, J.; Kishino, S.; Shimizu, S., CLA production from ricinoleic acid by lactic acid bacteria. *J. Am. Oil Chem. Soc.* **2003,** *80* (9), 889-894.

(67) Ando, A.; Ogawa, J.; Kishino, S.; Shimizu, S., Conjugated linoleic acid production from castor oil by *Lactobacillus plantarum* JCM 1551. *Enzyme Microb. Technol.* **2004,** *35* (1), 40-45.

(68) Rysman, E.; Brusselmans, K.; Scheys, K.; Timmermans, L.; Derua, R.; Munck, S.; Van Veldhoven, P. P.; Waltregny, D.; Daniëls, V. W.; Machiels, J., De novo lipogenesis protects cancer cells from free radicals and chemotherapeutics by promoting membrane lipid saturation. *Cancer Res.* **2010,** *70* (20), 8117-8126.

(69) Kennedy, S. G.; Wagner, A. J.; Conzen, S. D.; Jordan, J.; Bellacosa, A.; Tsichlis, P. N.; Hay, N., The PI 3-kinase/Akt signaling pathway delivers an anti-apoptotic signal. *Genes Dev.* **1997,** *11* (6), 701-713.

(70) Kulik, G.; Klippel, A.; Weber, M. J., Antiapoptotic signalling by the insulin-like growth factor I receptor, phosphatidylinositol 3-kinase, and Akt. *Mol. Cell. Biol.* **1997,** *17* (3), 1595-1606.

(71) Cuvillier, O., Sphingosine kinase-1–a potential therapeutic target in cancer. *Anti-Cancer Drugs* **2007,** *18* (2), 105-110.

(72) Illuzzi, G.; Bernacchioni, C.; Aureli, M.; Prioni, S.; Frera, G.; Donati, C.; Valsecchi, M.; Chigorno, V.; Bruni, P.; Sonnino, S., Sphingosine kinase mediates resistance to the synthetic retinoid N-(4-hydroxyphenyl) retinamide in human ovarian cancer cells. *J. Biol. Chem.* **2010,** *285* (24), 18594-18602.

(73) Hong, G.; Baudhuin, L. M.; Xu, Y., Sphingosine-1-phosphate modulates growth and adhesion of ovarian cancer cells. *FEBS Lett.* **1999,** *460* (3), 513-518.

(74) Bernstein, H.; Bernstein, C.; Payne, C.; Dvorakova, K.; Garewal, H., Bile acids as carcinogens in human gastrointestinal cancers. *Mutat. Res. - Rev. Mut. Res.* **2005,** *589* (1), 47-65.

(75) Temme, E. H.; Zhang, J.; Schouten, E. G.; Kesteloot, H., Serum bilirubin and 10-year mortality risk in a Belgian population. *Cancer Cause. Control* **2001,** *12* (10), 887-894.

(76) Stocker, R.; Yamamoto, Y.; McDonagh, A. F.; Glazer, A. N.; Ames, B. N., Bilirubin is an antioxidant of possible physiological importance. *Science* **1987,** *235* (4792), 1043-1046.

(77) Neužil, J.; Stocker, R., Bilirubin attenuates radical-mediated damage to serum albumin. *FEBS Lett.* **1993,** *331* (3), 281-284.

(78) Hussain, S. P.; Aguilar, F.; Amstad, P.; Cerutti, P., Oxy-radical induced mutagenesis of hotspot codons 248 and 249 of the human p53 gene. *Oncogene* **1994,** *9* (8), 2277-2281.

(79) Marks, J.; Davidoff, A.; Kerns, B.; Humphrey, P.; Pence, J.; Dodge, R.; Clarke-Pearson, D.; Iglehart, J.; Bast, R.; Berchuck, A., Overexpression and mutation of p53 in epithelial ovarian cancer. *Cancer Res.* **1991,** *51* (11), 2979-2984.

(80) Kupryjańczyk, J.; Thor, A. D.; Beauchamp, R.; Merritt, V.; Edgerton, S. M.; Bell, D. A.; Yandell, D. W., p53 gene mutations and protein accumulation in human ovarian cancer. *Proc. Natl. Acad. Sci. USA* **1993,** *90* (11), 4961-4965.

(81) Diamandis, E. P.; Yousef, G. M.; Soosaipillai, A. R.; Bunting, P., Human kallikrein 6 (zyme/protease M/neurosin): a new serum biomarker of ovarian carcinoma. *Clin. Biochem.* **2000,** *33* (7), 579-583.

(82) Tanimoto, H.; Yan, Y.; Clarke, J.; Korourian, S.; Shigemasa, K.; Parmley, T. H.; Parham, G. P.; O'Brien, T. J., Hepsin, a cell surface serine protease identified in hepatoma cells, is overexpressed in ovarian cancer. *Cancer Res.* **1997,** *57* (14), 2884-2887.

(83) Anisowicz, A.; Sotiropoulou, G.; Stenman, G.; Mok, S.; Sager, R., A novel protease homolog differentially expressed in breast and ovarian cancer. *Mol. Med.* **1996,** *2* (5), 624-636.

(84) Crowell, P. L.; Kennan, W. S.; Haag, J. D.; Ahmad, S.; Vedejs, E.; Gould, M. N., Chemoprevention of mammary carcinogenesis by hydroxylated derivatives of d-limonene. *Carcinogenesis* **1992,** *13* (7), 1261-1264.

(85) Ingram, D.; Sanders, K.; Kolybaba, M.; Lopez, D., Case-control study of phyto-oestrogens and breast cancer. *Lancet* **1997,** *350* (9083), 990-994.

(86) Scambia, G.; Ranelletti, F.; Panici, P. B.; Piantelli, M.; Bonanno, G.; De Vincenzo, R.; Ferrandina, G.; Rumi, C.; Larocca, L.; Mancuso, S., Inhibitory effect of quercetin on OVCA 433 cells and presence of type II oestrogen binding sites in primary ovarian tumours and cultured cells. *Br. J. Cancer* **1990,** *62* (6), 942.

(87) Ogretmen, B.; Hannun, Y. A., Biologically active sphingolipids in cancer pathogenesis and treatment. *Nat. Rev. Cancer* **2004,** *4* (8), 604-616.

(88) Senchenkov, A.; Litvak, D. A.; Cabot, M. C., Targeting Ceramide Metabolism—a Strategy for Overcoming Drug Resistance. *J. Natl. Cancer Inst.* **2001,** *93* (5), 347-357.

(89) Hannun, Y. A.; Luberto, C.; Argraves, K. M., Enzymes of sphingolipid metabolism: from modular to integrative signaling. *Biochemistry* **2001,** *40* (16), 4893-4903.

(90) Rylova, S.; Somova, O.; Dyatlovitskaya, E., Comparative investigation of sphingoid bases and fatty acids in ceramides and sphingomyelins from human ovarian malignant tumors and normal ovary. *Biochemistry-Moscow+* **1998,** *63* (9), 1057-1060.

(91) Lavie, Y.; Cao, H.-t.; Bursten, S. L.; Giuliano, A. E.; Cabot, M. C., Accumulation of glucosylceramides in multidrug-resistant cancer cells. *J. Biol. Chem.* **1996,** *271* (32), 19530-19536.

(92) Pettus, B. J.; Chalfant, C. E.; Hannun, Y. A., Ceramide in apoptosis: an overview and current perspectives. *BBA-Mol. Cell. Biol. L.* **2002,** *1585* (2–3), 114-125.

(93) Jones, C. M.; Monge, M. E.; Kim, J.; Matzuk, M. M.; Fernández, F. M., Metabolomic serum profiling detects early-stage high-grade serous ovarian cancer in a mouse model. *J. Proteome Res.* **2015,** *14* (2), 917-927.

(94) Vivanco, I.; Sawyers, C. L., The phosphatidylinositol 3-Kinase-AKT pathway in human cancer. *Nat. Rev. Cancer* **2002,** *2* (7), 489-501.

(95) Schöndorf, T.; Göhring, U. J.; Roth, G.; Middel, I.; Becker, M.; Moser, N.; Valter, M. M.; Hoopmann, M., Time to progression is dependent on the expression of the tumour suppressor PTEN in ovarian cancer patients. *Eur. J. Clin. Invest.* **2003,** *33* (3), 256-260.

(96) Kurose, K.; Zhou, X.-P.; Araki, T.; Cannistra, S. A.; Maher, E. R.; Eng, C., Frequent loss of PTEN expression is linked to elevated phosphorylated Akt levels, but not associated with p27 and cyclin D1 expression, in primary epithelial ovarian carcinomas. *Am. J. Pathol.* **2001,** *158* (6), 2097-2106.

(97) Wang, X.; Li, N.; Liu, B.; Sun, H.; Chen, T.; Li, H.; Qiu, J.; Zhang, L.; Wan, T.; Cao, X., A novel human phosphatidylethanolamine-binding protein resists tumor necrosis factor α-induced apoptosis by inhibiting mitogen-activated protein kinase pathway activation and phosphatidylethanolamine externalization. *J. Biol. Chem.* **2004,** *279* (44), 45855-45864.

(98) Li, P.; Wang, X.; Li, N.; Kong, H.; Guo, Z.; Liu, S.; Cao, X., Anti-apoptotic hPEBP4 silencing promotes TRAIL-induced apoptosis of human ovarian cancer cells by activating ERK and JNK pathways. *Int. J. Mol. Med.* **2006,** *18* (3), 505-510.

(99) Wang, X.; Li, N.; Li, H.; Liu, B.; Qiu, J.; Chen, T.; Cao, X., Silencing of human phosphatidylethanolamine-binding protein 4 sensitizes breast cancer cells to tumor necrosis factor-α–induced apoptosis and cell growth arrest. *Clin. Cancer. Res.* **2005,** *11* (20), 7545-7553.

(100) Fang, X.; Schummer, M.; Mao, M.; Yu, S.; Tabassam, F. H.; Swaby, R.; Hasegawa, Y.; Tanyi, J. L.; LaPushin, R.; Eder, A.; Jaffe, R.; Erickson, J.; Mills, G. B., Lysophosphatidic acid is a bioactive mediator in ovarian cancer. *BBA-Mol. Cell. Biol. L.* **2002,** *1582* (1–3), 257-264.

(101) Zwaal, R. F. A.; Schroit, A. J., *Pathophysiologic Implications of Membrane Phospholipid Asymmetry in Blood Cells*. 1997; Vol. 89, p 1121-1132.

(102) Martin, S. J.; Reutelingsperger, C. P.; McGahon, A. J.; Rader, J. A.; van Schie, R. C.; LaFace, D. M.; Green, D. R., Early redistribution of plasma membrane phosphatidylserine is a general feature of apoptosis regardless of the initiating stimulus: inhibition by overexpression of Bcl-2 and Abl. *J. Exp. Med.* **1995,** *182* (5), 1545-1556.

(103) Rose, D. P.; Connolly, J. M., Antiangiogenicity of docosahexaenoic acid and its role in the suppression of breast cancer cell growth in nude mice. *Int. J. Oncol.* **1999,** *15* (5), 1011-1016.

(104) Larsson, S. C.; Kumlin, M.; Ingelman-Sundberg, M.; Wolk, A., Dietary long-chain n−3 fatty acids for the prevention of cancer: a review of potential mechanisms. *Am. J. Clin. Nutr.* **2004,** *79* (6), 935-945.

(105) Pidgeon, G. P.; Kandouz, M.; Meram, A.; Honn, K. V., Mechanisms controlling cell cycle arrest and induction of apoptosis after 12-Lipoxygenase inhibition in prostate cancer cells. *Cancer Res.* **2002,** *62* (9), 2721-2727.

(106) Ding, X.-Z.; Tong, W.-G.; Adrian, T. E., 12-lipoxygenase metabolite 12(S)-HETE stimulates human pancreatic cancer cell proliferation via protein tyrosine phosphorylation and ERK activation. *Int. J. Cancer* **2001,** *94* (5), 630-636.

(107) Rose, D. P.; Connolly, J. M., Regulation of Tumor Angiogenesis by Dietary Fatty Acids and Eicosanoids. *Nutr. Cancer* **2000,** *37* (2), 119-127.

(108) Berger, J.; Moller, D. E., The mechanisms of action of PPARS. *Annu. Rev. Med.* **2002,** *53* (1), 409-435.

(109) Eggens, I.; Elmberger, P.; Löw, P., Polyisoprenoid, cholesterol and ubiquinone levels in human hepatocellular carcinomas. *Br. J. Exp. Pathol.* **1989,** *70* (1), 83.

(110) Olsson, J. M.; Schedin, S.; Teclebrhan, H.; Eriksson, L. C.; Dallner, G., Enzymes of the mevalonate pathway in rat liver nodules induced by 2-acetylaminofluorene treatment. *Carcinogenesis* **1995,** *16* (3), 599-605.

(111) Rosen, D. G.; Yang, G.; Liu, G.; Mercado-Uribe, I.; Chang, B.; Xiao, X. S.; Zheng, J.; Xue, F.-X.; Liu, J., Ovarian cancer: pathology, biology, and disease models. *Front. Biosci.* **2009,** *14*, 2089-2102.

(112) Feeley, K.; Wells, M., Precursor lesions of ovarian epithelial malignancy. *Histopathology* **2001,** *38* (2), 87-95.

(113) Olivier, R.; Van Beurden, M.; Van'T Veer, L., The role of gene expression profiling in the clinical management of ovarian cancer. *Eur. J. Cancer* **2006,** *42* (17), 2930-2938.

PART II: CHALLENGES IN MASS SPECTROMETRY-BASED UNTARGETED METABOLOMICS

CHAPTER 4. METABOLOMICS AND PROTEOMICS REVEAL METABOLIC IMPACTS OF CHEMICALLY MEDIATED COMPETITION ON MARINE PLANKTON

Adapted with permission from
Jones, C. M. ‡; Ellestad, K. L. ‡; Roy, J.; Viant, M. R.; Fernández, F. M.; Kubanek, J.; Nunn, B. L., Metabolomics and Proteomics Reveal Impacts of Chemically Mediated Competition on Marine Plankton. *Proc. Natl. Acad. Sci. USA* **2014**, *111* (24), 9009-9014.
‡equal contributing author

This chapter describes research conducted by multiple persons. K. L. Poulson-Ellestad conducted all NMR metabolomics experiments and analyzed and interpreted the resultant data in addition to conducting the phytoplankton culturing experiments. C. M. Jones assisted with phytoplankton harvesting and metabolite extraction in addition to conducting all MS metabolomics experiments and analyzing and interpreting the resultant data. B. L. Nunn conducted all MS proteomics experiments and analyzed and interpreted the resultant data.

4.1 Abstract

Karenia brevis is a toxic dinoflagellate known to exude allelopathic compounds that directly inhibit the growth of species with whom it competes for resources. Yet, brevetoxins produced by *K. brevis*, which are responsible for shellfish and marine mammal deaths, are known not to be allelopathic. *K. brevis* allelopathic compounds have not yet been identified, and their metabolic effects are poorly understood. In this work, the combined power of metabolomics and proteomics was exploited to explore how chemically mediated interactions between *K. brevis* and two diatom competitors, *Asterionellopsis glacialis* and *Thalassiosira pseudonana*, impact competitor physiology. This integrated systems biology showed that *K. brevis* allelopathy distinctively perturbed the metabolisms of these two competitors. *A. glacialis* had a more robust metabolic response to *K. brevis* allelopathy which may be a result of its repeated exposure to *K.*

brevis blooms in the Gulf of Mexico. However, *K. brevis* allelopathy disrupted energy metabolism and obstructed cellular protection mechanisms including altering cell membrane components, inhibiting osmoregulation, and increasing oxidative stress in *T. pseudonana.* This work highlights the ability of systems biology to shed light onto the nature of complex ecological interactions.

4.2 Marine Plankton Competition

4.2.1 Chemically Mediated Competition

Fifty percent of net primary production is generated by marine phytoplankton.[1] Higher trophic levels require primary production for survival; however, interactions among marine planktonic species affect the movement of the produced biomass and nutrients in the water column.[2] Numerous interactions among marine planktonic species, such as competition[3, 4] and the capture of prey,[5] are mediated by chemical signals and cues. Consequently, chemically mediated ecological interactions may impact large-scale ecosystem processes.

Allelopathy, an interference competition method through which species release biochemical compounds that kill or inhibit the growth of competitors, is capable of modifying the composition of terrestrial[6] and benthic aquatic communities.[7, 8] Allelopathic interactions among marine planktonic species can cause immense competitor death[9] or induce apoptosis.[10] Conversely, allelopathy can also cause sublethal effects, such as reduced growth rate[11, 12] or altered cell swimming behavior.[4, 13]

4.2.2 *Karenia brevis*

Karenia brevis is a red tide dinoflagellate that exudes brevetoxins during red tide blooms in the Gulf of Mexico. These brevetoxins cause fish and marine mammal death in addition to neurotoxic shellfish poisoning in humans.[14] Although brevetoxins are lethal, they are known not to be allelopathic.[3, 15-17] *K. brevis* also exudes a set of polar biochemical compounds[16] of unknown identity that harmfully impacts the physiology of certain competitor phytoplankton species.[3, 17] These unidentified allelopathic compounds inhibit competitor growth and photosynthetic processes as well as cause increased cell membrane permeability.[17] Interestingly, competitor species are differentially affected by *K. brevis* allelopathy and the degree to which competitors are affected is in part facilitated by ecological context.[3] Still, the cellular targets of *K. brevis* allelopathy remain unknown.

4.2.3 Omics Integration

The aim of systems biology is to reconstruct cellular network interactions to predict and characterize physiological responses to stimuli. Data from systems biology fields must be combined to accomplish such a vast task. It has been suggested that omics integration has the potential to identify basic concepts of cellular function that are currently unknown.[18] As proteins are responsible for cellular signals, structural integrity, and catalysis of most biochemical reactions including the production and conversion of the vast array of metabolites required for cellular survival, integrating findings from metabolomic and proteomic analyses gives more global insight into cellular behavior. Although metabolomics and proteomics have been used in the past to examine diatoms adapting to various stressors,[19-22] the presented work represents the first instance of

metabolites and proteins measured simultaneously to understand the effects of allelopathy or in fact any form of competition.

4.3 Experimental Methods

Experimental details for phytoplankton culturing and monitoring, the experimental co-culture design and preparation, nutrient analyses, harvest, extraction, and processing of phytoplankton cultures, and NMR spectral acquisition and processing can be found elsewhere.[23]

4.3.1 Ultra Performance Liquid Chromatography-Mass Spectrometry Data Acquisition

Prior to UPLC-MS analysis, all diatom polar extracts were reconstituted in 250 µL of H_2O. UPLC-MS analysis was performed using a Waters Xevo G2 mass spectrometer (Waters Corporation, Manchester, UK) coupled to a Waters ACQUITY UPLC H Class system, fitted with a Waters ACQUITY UPLC BEH C18 column (2.1 × 50 mm, 1.7 µm particle size; Waters Corporation, Milford, MA, USA). The Xevo G2 mass spectrometer has a characteristic resolving power of 25,000 M/Δ*m* (FWHM) and mass accuracy of 1.8 ppm at m/z 554.2615. The instrument was operated in negative electrospray ionization mode with a probe capillary voltage of 2.5 kV, and a sampling cone voltage of 45 V. The source and desolvation nitrogen gas temperatures were 120 °C and 350 °C, respectively. The nitrogen desolvation flow rate was 800 L h^{-1}. The mass spectrometer was calibrated across the 50-1200 Da mass range using a 0.5 mM sodium formate solution prepared in 90:10 2-propanol/water (v/v). A leucine enkephalin reference spray (LockSpray) infused at 3 µL min^{-1} was utilized for mass correction

during data acquisition. Data were acquired in the 50-1200 m/z range with a scan time of 1 s. Data acquisition and processing was carried out using MassLynx v4.1. UPLC-MS/MS experiments were performed by acquiring mass spectra with applied voltages of 10, 20, 30 and 40 V in the T-wave collision cell using ultra high purity argon (≥ 99.999%) as the collision gas.

Chromatographic separation solvents for sample analysis were H_2O with 0.1 % acetic acid (mobile phase A) and acetonitrile (mobile phase B). The following mobile phase gradient was utilized: 0–1 min, 0% B; 1–3 min, 0–61% B; 3–5 min, 61% B; 5–6 min, 61–62% B; 6–8 min, 62% B; 8–10 min, 62–100% B; 10–15 min, 100% B. The flow rate was held constant at 0.3 mL min^{-1}. The injection volume was 7 µL. The chromatographic method for re-equilibrating the column after each sample injection was as follows: 0–1 min, 0% A; 1–1.5 min, 0–100% A; 1.5–2.5 min, 100–0% A; 2.5–3.5 min, 0–100% A; 3.5–5 min, 100% A. The flow rate was held constant at 0.3 mL min^{-1} during the re-equilibration period. The column and autosampler tray temperatures were 35 °C and 5 °C, respectively.

4.3.2 Ultra Performance Liquid Chromatography-Mass Spectrometry Data Processing

Following UPLC-MS data acquisition, metabolic features (retention time, m/z pairs) were extracted from chromatograms using MarkerLynx XS software (Waters Corporation, Manchester, UK). This procedure involved chromatogram alignment, peak picking and integration, peak area extraction, and normalization of all data. Technical duplicate data were acquired for all samples and the resulting feature peak areas were averaged to obtain the finalized peak area list for each biological replicate. Metabolic

features detected in *K. brevis* blank culture extracts in addition to media blank culture extracts were excluded from the feature list.

4.3.3 Metabolomics Statistical Analyses

PCA was used to initially examine differences in diatom metabolomes between treatment and controls of the same species (MATLAB with PLS Toolbox v.7.03). To determine statistical differences between PCA scores of treatment and control extracts, t tests were performed (MATLAB). PCA models for both the *T. pseudonana* and *A. glacialis* experiments are provided in Figure 4.3, but these were only used as an exploratory tool. PLS-DA data models used to investigate metabolome differences for both experiments are shown in Figure 4.2. UPLC-MS data were preprocessed by autoscaling prior to statistical analyses. PLS-DA data models were orthogonalized and internally cross-validated using the venetian blinds method with 7 data splits in an effort to extract all relevant class-discriminatory information from the first latent variable, ensure accuracy, and assess for any overfitting.

For NMR data, after PLS-DA, the weights on latent variable 1 were used to ascertain which chemical shifts (i.e., metabolites) within extracts were responsible for distinguishing between *K. brevis* exposed and media exposed cultures. For UPLC-MS data, forward interval PLSDA (iPLSDA) variable selection was used to enhance the accuracy of the PLS-DA models and simplify their complexity. The iPLSDA number of intervals and step size were chosen automatically, while the interval size was set to one, and the maximum number of latent variables was ten. Additionally, for the *T. pseudonana* experiment, selectivity ratios (the ratio of a feature's explained and residual variance)[24] were employed to assist in the determination of useful metabolic features that

had discriminatory power between *K. brevis*-exposed and media-exposed species. Forward iPLSDA variable selection only provided one metabolic feature (2.68 min, m/z 701.0880) responsible for discrimination, so selectivity ratios were used to extract more informative metabolic information from the *T. pseudonana* data set. Metabolic features with selectivity ratios greater than 1.5 were used. However, selectivity ratios were not utilized for the *A. glacialis* experiment since there were no metabolic features with selectivity ratios greater than one.

As the biological hypotheses presented in this work were based on multivariate analysis via PLS-DA, changes in concentration of single metabolites were combined into a discriminating panel that was used for the separation of *K. brevis*-exposed and media-exposed cultures. P-values are noted in Table 4.2 to present a view of the univariate difference in metabolite concentrations, even if the univariate difference alone was not significant ($p \leq 0.05$).

4.3.4 Metabolite Annotation

Chemical identification was attempted for all discriminant features. Representative extracts from each experiment were used to collect 2D NMR spectral data (standard HSQC and TOCSY experiments, 500 MHz) to aid in annotation. The Kyoto Encyclopedia of Genes and Genomes (KEGG)[25] was also used to aid in identification of particular pathways involving critical metabolites. For UPLC-MS, because numerous adduct ions can be detected for a single compound with electrospray ionization, adduct ion analysis was initially conducted to ascertain correct metabolite identifications. The adduct ions that were investigated in the mass spectra included $[M-H]^-$, $[M+Cl]^-$, $[M+Br]^-$, $[M+ CH_3COO]^-$, $[M+HCOO]^-$, $[M+Na-2H]^-$, $[M-H_2O-H]^-$ and $[2M-H]^-$ species, which

are commonly observed in negative electrospray ionization mode. The theoretical m/z values for typical adduct species were calculated and compared with the experimental values from mass spectral signals. For spectra in which multiple adducts were not present, the accurate mass of the candidate neutral molecule was calculated based on the assumption that the m/z value for the metabolic feature of interest corresponded to $[M-H]^-$. For mass spectra in which multiple adducts were present, the $[M-H]^-$ spectral signal was identified and the accurate mass of the candidate neutral molecule was calculated based on it. Molecular formulae were generated based on the mass accuracy of the discriminant metabolic features of interest, in addition to their isotopic patterns, using MassLynx 4.1.

The lists of generated molecular formulae and chemical shifts were searched against the METLIN database,[26] KEGG,[25] LIPID Metabolites and Pathways Strategy (LIPID MAPS) database,[27] the Madison-Qingdao Metabolomics Consortium Database (MMCD),[28] the Human Metabolome Database (HMDB),[29] and Chenomx Profiler, as well as literature searches used to annotate spectra and tentatively identify metabolites. MetaboSearch[30] was also utilized to search the aforementioned databases solely using neutral masses with a mass accuracy of 20 ppm. Tandem MS data could not be acquired for all discriminant features as their ion abundances were not intense enough for sensitive quadrupole mass selection due to ion transmission losses.[31] Acquired fragmentation patterns from MS/MS data could not be utilized for further identification due to the co-elution of compounds with the features of interest. Some of the discriminatory metabolic features were matched to metabolites specific to plants. In such cases, the metabolites were described by compound class, and are included in the metabolite identification

tables in italics (Tables 4.3 and 4.5). As numerous compounds are still being discovered in microalgae,[32] it is plausible that isomers of known specific plant metabolites could exist in diatoms.

4.3.5 Protein Sample Preparation

Two biological replicate samples from each *T. pseudonana* treatment and three biological replicates from each *A. glacialis* treatment were collected for proteomic analyses (450 ml culture each). Cells were pelleted (10,000 *g*; 10 min) on ice and lysed using a titanium microtip sonicating probe. Each sample received 10 sonication events (10-15 s each) in 0.2% PPS silent surfactant[33] (Agilent Technologies) in 50 mM ammonium bicarbonate. The details of the digestion were per the manufacturer's guidelines. Disulfide bonds were reduced with dithiothreitol and alkylated with iodoacetamide. Each sample received trypsin at an enzyme to protein ratio of 1:50, vortexed, and incubated on a Thermomixer (800 rpm) 4 h at 37 ^{o}C. Peptide concentrations were measured for each sample using a Thermo Scientific NanoDrop 2000/2000c spectrophotometer. The peptide bond absorbance was monitored at 205 nm and samples were diluted to yield a final concentration of 100 μg protein mL^{-1}.

4.3.6 Mass Spectrometry-Based Proteomics

Samples were separated and introduced into the mass spectrometer by reversed-phase chromatography using a 30 cm-long, 75 μm i.d. fused silica capillary column packed with C_{18} silica particles (Magic C18AQ, 100 Å, 5 m; Michrom, Bioresources, Inc., CA) fitted with a 2 cm long, 100 μm i.d. precolumn (Magic C18AQ, 200 Å, 5 m; Michrom).[19] Peptides were eluted using an acidified (formic acid, 0.1% v/v)

water/acetonitrile gradient (2-35% acetonitrile over 90 min). Mass spectrometry was performed on a Thermo Fisher (San Jose, CA) QExactive (QE). Based on peptide concentrations, a total of 1 µg of peptide digest in 10 µl of 2% ACN, 0.1% formic acid was sampled per LC-MS analysis. Biological duplicates were analyzed from the treatment and the control group of *T. pseudonana* using data-dependent acquisition (DDA). Biological triplicates of *A. glacialis* controls and treatment were analyzed using DDA on the QExactive. To determine relative expression of proteins with respect to the alternate cell state, *QSpec* software was utilized to provide statistical significance.[34]

4.3.7 Protein Database Searching and Mass Spectrometry Data Interpretation

Tandem mass spectrometry results were searched and interpreted with SEQUEST (PVM v.27 20070905). SEQUEST parameters included: reverse concatenated sequence database search, trypsin enzyme specificity, cysteine modification of 57 Da (resulting from iodoacetamide), and modifications on methionine of 15.999 Da (oxidation). Minimum protein and peptide thresholds were set at $p > 0.95$ on ProteinProphet and PeptideProphet.[35, 36] Protein identifications from whole cell lysates were accepted by ProteinProphet if the above mentioned thresholds were passed, two or more peptides were identified (PeptideProphet), and at least one terminus was tryptic. Using concatenated target-decoy database searches, false discovery rates (FDR) were calculated according to Elias and Gygi 2007[37] and were all < 1%. The protein database used for correlating spectra from the *T. pseudonana* experiment with protein identification was generated by combining the latest release of the *T. pseudonana* proteome (12914 proteins) and 50 common contaminants. Since the genome of *A. glacialis* has not been described, we first searched the *A. glacialis* data as for *T. pseudonana*. Second, we

assembled a database to search for confident peptide matches to spectra from protein sequences associated with diatoms in the National Center for Biotechnology Information database (www.ncbi.nlm.nih.gov). Using the "nrdb90" script provided by Holm and Sander [38] at a 95% identity threshold on the input diatom database of 93,744 protein sequence entries reduced redundancy and size down to 64,752 sequence entries.[39] ProteinProphet and PeptideProphet were set at a higher threshold of $p > 0.99$, to increase confidence in peptide and protein assignments. Proteins from the *A.glacialis* treatments and controls were considered present if one of the biological replicates identified two or more unique peptides and the termini were tryptic.

To determine relative expression of proteins expressed in treatments with respect to the control cell state, *QSpec* software was utilized to provide statistical significance.[34] The Database for Annotation, Visualization and Integrated Discovery (DAVID) v. 6.7 was used to examine enriched biological terms in the sets of proteins identified to be significantly enhanced or suppressed relative to control cell state.[40, 41] P-values <1E-3 are reported for DAVID analyses on *T. pseudonana* data.

4.4 *Karenia brevis* Allelopathy Impacts Competitor Metabolism

Although previous studies[3, 17] have shown that *K. brevis* allelopathy reduces the growth of competitor species, the current study demonstrates the impact of allelopathy on global cellular function and indicates that metabolism was altered even in the case of modest growth inhibition.

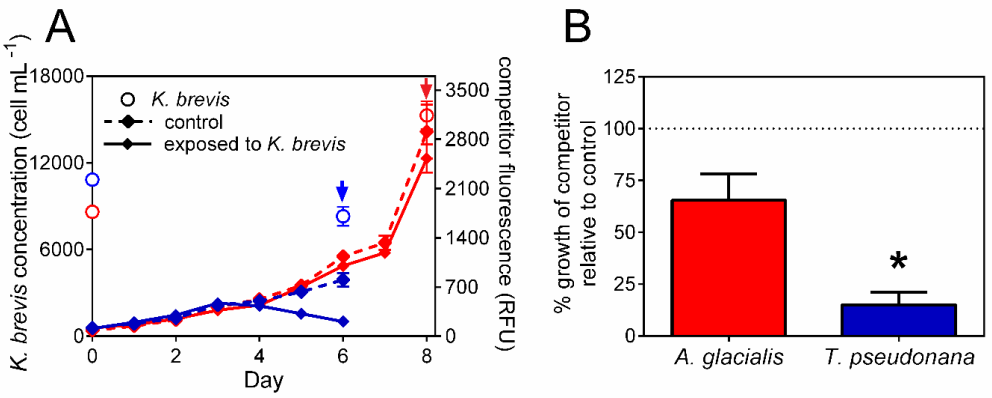

Figure 4.1: Effects of exposure to exudates of live *Karenia brevis* on the growth of *Asterionellopsis glacialis* and *Thalassiosira pseudonana.* (A) *A. glacialis* (red) *in vivo* and *T. pseudonana* (blue) *in vivo* fluorescence (arrow indicates day of harvest for metabolomics and proteomics). The solid lines indicate fluorescence of diatom-only controls, and the dashed lines indicate fluorescence of diatoms exposed to *K. brevis*. Initial *K. brevis* (red open circles for *A. glacialis* experiment; blue open circles for *T. pseudonana* experiment) concentrations from cultures used to fill dialysis tubes (n = 1), final concentrations from experimental flasks at time of harvest (n = 15). (B) Calculated percent growth of competitors *A. glacialis* (red) and *T. pseudonana* (blue) relative to their own controls after 8 and 6 days exposure to *K. brevis*, respectively. The dotted line indicates growth equivalent to control. n = 15. $P < 0.0001$ indicated by asterisk (*), unpaired t-test. Error bars represent ±1 S.E.M.

K. brevis allelopathy differentially affected growth of the two competitors (Figure 4.1), which was reflected in both the metabolomes (Figures 4.2, 4.3) and proteomes (Table 4.1) of each competitor. In the presence of *K. brevis*, *T. pseudonana* growth was reduced by 85% (Figure 4.1b, n = 15, $p < 0.0001$). Growth suppression and subsequent population decline was observed after four days of co-culture with both species physically separated by dialysis membrane.

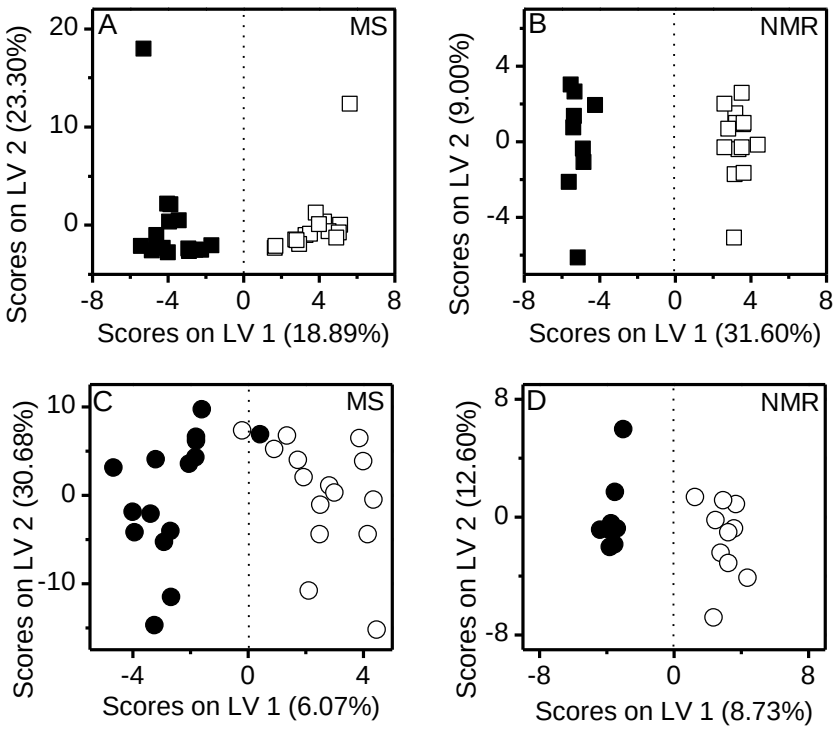

Figure 4.2: Orthogonal projection to latent structures-discriminant analysis (oPLS-DA) shows effects of *Karenia brevis* allelopathy on the metabolomes of competitor diatoms. oPLS-DA calibration scores plot of (A) UPLC-MS metabolic features and (B) ^{1}H NMR spectral data for *Thalassiosira pseudonana* exposed to *K. brevis* (filled squares) or dilute media control (empty squares) with cross-validated accuracies of 87% and 100%, respectively. oPLS-DA calibration scores plot of (C) UPLC-MS metabolic features and (D) ^{1}H NMR spectral data for *Asterionellopsis glacialis* exposed to live *K. brevis* (filled circles) or dilute media controls (empty circles) with cross-validated accuracies of 57% and 63%, respectively.

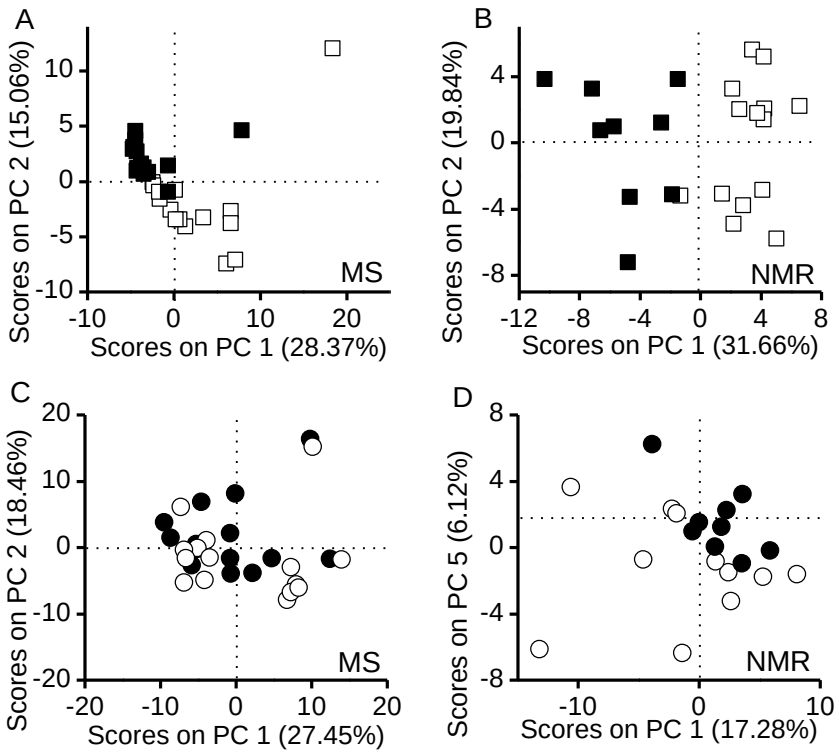

Figure 4.3: Principal component analysis (PCA) shows effects of *Karenia brevis* allelopathy on the metabolomes of competitor diatoms. PCA scores plot of (A) UPLC-MS metabolic features and (B) ^{1}H NMR spectral data for *Thalassiosira pseudonana* exposed to *K. brevis* (filled squares) or dilute media control (empty squares) showing significant separation along the 2nd and 1st principal components, respectively (MS: unpaired t-test, n = 30, $p = 0.002$; NMR: unpaired t-test, n = 9-14, $p < 0.0001$). PCA scores plot of (C) UPLC-MS metabolic features and (D) ^{1}H NMR spectral data for *Asterionellopsis glacialis* exposed to live *K. brevis* (filled circles) or dilute media controls (empty circles) with only the ^{1}H NMR spectral data showing significant separation (5th principal component, unpaired t-test, n = 9-11, $p = 0.033$).

Analysis of the *T. pseudonana* metabolome (Table 4.2; Figure 4.2) and proteome (Table 4.1) revealed that *K. brevis* allelopathy impacts multiple metabolic pathways, with some cellular responses possibly used to compensate for negative effects of allelopathy. In contrast, the overall growth of *A. glacialis* was only reduced by 35% when exposed to *K. brevis* allelopathy—an effect that was not statistically significant (Figure 4.1b, n = 15, $p = 0.089$) and, at most, indicative of a weak allelopathic effect of *K. brevis* on *A. glacialis* in this experiment. In the *A. glacialis* experiment, final *K. brevis* cell concentrations in flasks averaged 15.3 x 10^3 cell mL^{-1}, whereas mean *K. brevis* concentrations were 8.3 x 10^3 cell mL^{-1} at the time of harvest in the *T. pseudonana* experiment (Figures 4.1a). Nutrients were not limiting in these experiments, suggesting that the negative effects of *K. brevis* on competitor growth were mediated by allelopathy rather than exploitation competition. Overall, it is clear that based on both the proteomic and metabolomic data, a sensitive competitor exposed to allelopathy operates in a functionally compromised metabolic state (Figures 4.2, 4.3).

Table 4.1: Metabolic pathways and cellular functions from Biological Process and Molecular Function Gene Ontology categories for proteins in *Thalassiosira pseudonana* whose concentrations increased (red triangle) or decreased (blue triangle) in response to *Karenia brevis* allelopathy.

	Metabolic Pathway/Function	# of Proteins[1]	*p* value[2]
▲	Generation of precursor metabolites and energy	25	8.9E-12
▲	Metabolic process	101	3.2E-10
▲	Cellular carbohydrate metabolic process	16	2.0E-08
▲	Monosaccharide metabolic process (hexose and glucose)	14	2.1E-08
▲	Alcohol metabolic process	15	3.6E-08
▲	Energy coupled proton transport, down electrochemical gradient	10	4.7E-07
▲	ATP metabolic process	11	1.7E-06
▲	Oxidative phosphorylation	10	3.8E-06
▲	Nucleoside triphosphate metabolic process	11	4.5E-06
▲	Cellular metabolic process	73	5.7E-06
▲	Proton-transporting ATPase activity, rotational mechanism	9	9.7E-10
▲	H-ion transport ATP synthase activity, rotational mechanism	8	2.2E-08
▲	Catalytic activity	96	4.2E-08
▲	ATPase activity, coupled to transmembrane movement of ions	10	9.3E-08
▲	Hydrolase activity, acting on acid anhydrides, catalyzing transmembrane movement of substances	11	8.2E-07
▲	Lyase activity	15	2.2E-06
▲	Cofactor binding	19	3.3E-06
▼	Photosynthesis	12	7.7E-11
▼	Generation of precursor metabolites and energy	13	1.3E-09
▼	Chromatin assembly	8	1.6E-09
▼	Cellular macromolecular complex subunit organization	8	3.5E-07
▼	Cellular component biogenesis	9	8.5E-07
▼	Electron carrier activity	6	1.3E-03
▼	Iron ion binding	6	2.8E-03
▼	4 Iron, 4 sulfur cluster binding	3	3.0E-03
▼	Metal cluster binding	4	4.3E-03

[1]# of proteins represents the number of proteins that correlate to the annotation category.
[2]*P* value is the modified Fisher Exact p value for protein enrichment analysis.

Table 4.2: Candidate metabolites tentatively identified by NMR spectra and UPLC-MS to increase in concentration (red arrows) or decrease in concentration (blue arrows) in *Thalassiosira pseudonana* exposed to *Karenia brevis*.

Biological Category	**Candidate Metabolite**	**Biological Sub-category**	**Fold Change[1]**	**Method of Detection**	***p* value[2]**
Cell Protection					
▲	Lipid ($C_{31}H_{42}O_9$)[3]	Membrane constituent	+10.3	UHPLC/MS	6.9E-17
▲	Phospholipid ($C_{24}H_{47}O_{12}P$)[3]	Membrane constituent	+10.3	UHPLC/MS	6.9E-17
▲	Sarcosine[4]	Glycine metabolism/Osmoregulation	+0.32	^{1}H-HSQC NMR	4.9E-1
▼	Mannan[5,6]	Cell wall constituent	-3.4	UHPLC/MS	9.8E-16
▼	Phospholipid ($C_{26}H_{43}O_9P$)[7]	Membrane constituent	-1.4	UHPLC/MS	2.5E-15
▼	Lipid ($C_{26}H_{42}O_{11}$)[7]	Membrane constituent	-1.4	UHPLC/MS	2.5E-15
▼	Terpene glycoside	Membrane constituent	-1.2	UHPLC/MS	6.9E-15
▼	Betaine	Osmoregulation	-0.58	^{1}H-HSQC NMR	2.9E-3
▼	Myo-inositol[8]	Osmoregulation	-0.41	^{1}H-HSQC NMR	6.2E-1
▼	Homarine	Osmoregulation	-2.5	^{1}H-HSQC NMR	3.0E-5
▼	Polyphenol[5]	Oxidative stress	-3.4	UHPLC/MS	9.8E-16
▼	Taurine	Amino Acid metabolism	N/A	^{1}H-HSQC NMR	N/A

Table 4.2 (continued).

Energy Metabolism					
▲	Acetate	Carbon Metabolism/Glycolysis	+0.92	^{1}H-HSQC NMR	1.8E-2
▲	Dimethylamine[4]	Glycine metabolism/Osmoregulation	+0.32	^{1}H-HSQC NMR	4.9E-1
▼	Alanine	Amino Acid metabolism	-0.48	^{1}H-HSQC NMR	1.1E-1
▼	Stachyose[5,6]	Carbon metabolism	-3.4	UHPLC/MS	9.8E-16
▼	Cellotetraose[5,6]	Carbon metabolism	-3.4	UHPLC/MS	9.8E-16
▼	Maltopentaose[9]	Carbon metabolism	-3.2	UHPLC/MS	3.3E-20
▼	Amylopectin[9]	Carbon metabolism	-3.2	UHPLC/MS	3.3E-20
▼	Cellopentaose[9]	Carbon metabolism	-3.2	UHPLC/MS	3.3E-20
▼	Verbascose[9]	Carbon metabolism	-3.2	UHPLC/MS	3.3E-20
▼	Glucose[8]	Carbon Metabolism/Glycolysis	-0.41	^{1}H-HSQC NMR	2.1E-1
▼	Glycerate	Photorespiration	-0.024	^{1}H-HSQC NMR	9.2E-1
▼	Dihydrouracil	Pyrimidine metabolism	-0.91	^{1}H-HSQC NMR	2.4E-2
Cell Protection/Energy Metabolism					
▲	Glutamate	Amino Acid metabolism/ osmoregulation	+1.22	^{1}H-HSQC NMR	1.4E-3

Table 4.2 (continued).

▲	Proline	Amino Acid metabolism/ osmoregulation	+0.052	[1]H-HSQC NMR	8.8E-1

[1]Fold change was calculated as the base 2 logarithm of the average peak area ratios for exposed *T. pseudonana* samples and control samples; [2]*P* values calculated using unpaired, two-tailed *t*-test; [3]isobars; [4]metabolites share observed chemical shifts; [5]isobars; [6]isomers; [7]isobars; [8]metabolites share multiple observed chemical shifts; [9]isomers

4.5 Response of *Asterionellopsis glacialis* to Allelopathy

To date, metabolomics and proteomics are under-utilized tools for investigating the impacts of ecological interactions on organism physiology. This study revealed that metabolomic and proteomic analyses of competitors can demonstrate the species-specific effects of *K. brevis* allelopathy. The diatom *A. glacialis* appeared to maintain a relatively robust metabolism in response to *K. brevis,* which resulted in mild allelopathic effects on growth (Figure 4.1, Figure 4.2c and 4.2d). Using multivariate analysis [orthogonal projection to latent structures-discriminant analysis (oPLS-DA)], we identified groups of metabolites whose co-varying concentrations allowed us to distinguish metabolomes of *K. brevis*-exposed and control cultures of *A. glacialis*. While concentrations of these compounds may not have significantly differed between treatments and controls when each compound was considered alone, these compounds displayed explanatory power when part of a discriminating panel.[42] oPLS-DA analysis suggested that concentrations of only a few aliphatic metabolites were enhanced in *A. glacialis* by exposure to allelopathy (i.e., glycerophosphocholine and lactose or polysaccharides, as well as two unidentified compounds; Table 4.3). However, only 6-9% (Figure 4.2c and 4.2d) of the observed among-culture variation in the *A. glacialis* metabolome was due to *K. brevis* allelopathy, indicating that this species maintains relatively normal functioning in response to *K. brevis*. Additionally, only one protein increased in abundance when exposed to *K. brevis* allelopathy according to differential expression protein analyses (ATP synthase) whereas five proteins were significantly suppressed (e.g., cytochrome b559, photosystem II chlorophyll A antenna apoprotein; Dataset 4.1).

Table 4.3: Candidate metabolites tentatively identified from UPLC-MS data of *Asterionellopsis glacialis* exposed to *Karenia brevis* allelopathy.

Retention Time (min)	m/*z*	Metabolite Candidates in Database	Molecular Formula	Mass Error (mDa)	Detected Ion Type	Effect of *K. brevis*	Metabolic Pathway/Function
5.31	1011.564	*Amido disaccharide*	$C_{22}H_{40}N_2O_{13}$	5.5	$[M+Cl]^-$	Enhanced	-
4.60	539.2513	*Steroid glycoside*	$C_{27}H_{40}O_{11}$	1.5	$[M-H]^-$	Suppressed	Membrane stability; synthesis of secondary metabolites [43]

italics indicate compounds not previously described from marine planktonic organisms

The strain of *A. glacialis* used in this experiment was originally isolated from the Gulf of Mexico, where *K. brevis* blooms regularly occur.[14] Thus, *A. glacialis* may have evolved partial resistance to *K. brevis* allelopathy due to its potential for frequent exposure to *K. brevis* blooms.[44] Conversely, *T. pseudonana* used in the current study was from the North Atlantic Ocean where *K. brevis* does not occur. It is possible that the robust metabolome of *A. glacialis* has resulted from multiple metabolic adaptations that mitigate allelopathic impacts, allowing *A. glacialis* cells to function when challenged by *K. brevis* allelopathy.

4.6 Response of *Thalassiosira pseudonana* to Allelopathy

Because of its sensitivity to *K. brevis* allelopathy (Figure 4.1), *T. pseudonana* serves as a suitable model for exploring mechanisms by which allelopathy affects competitor physiology. From our integrated systems biology approach we found *T. pseudonana* energy metabolism to be greatly impacted by *K. brevis* allelopathy (Figure 4.4).

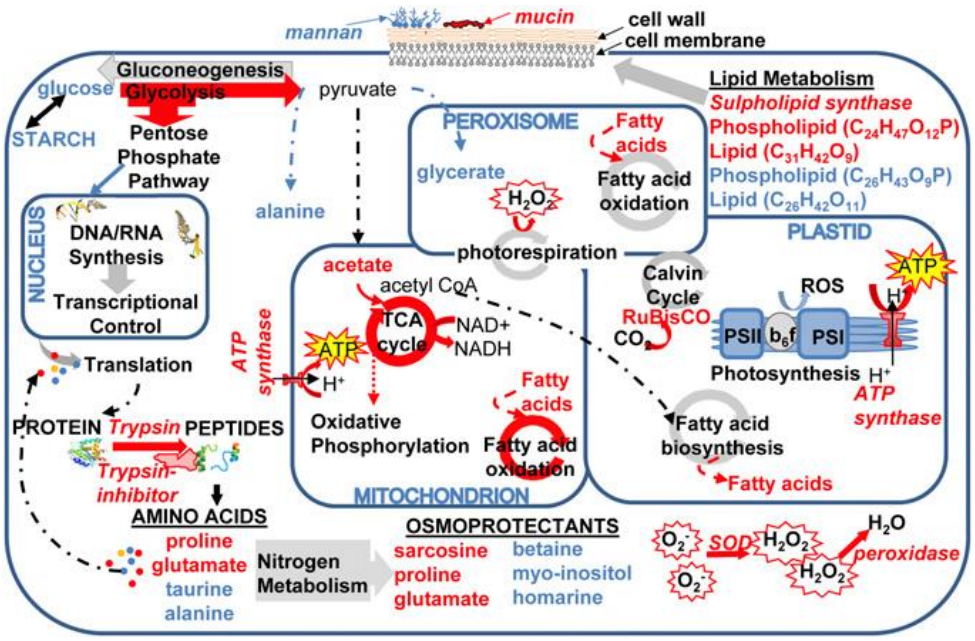

Figure 4.4: Network of cellular pathways, enzymes, and metabolites in the diatom *Thalassiosira pseudonana* impacted by exposure to *Karenia brevis* allelopathy, derived from NMR and MS metabolomics and proteomics. Pathways and metabolites enhanced by allelopathy are indicated by red arrows and compound names, respectively. Blue arrows and compound names denote pathways and metabolites that were suppressed by allelopathy.

An observed significant increase of acetate concentrations (+0.92-fold), as well as decreased *myo*-inositol and/or glucose concentrations, as distinguished by oPLS-DA (Table 4.2, Table 4.4), with concomitant increased concentrations of glycolysis enzymes (e.g., pyruvate kinase +3.0-fold; Table 4.1, Dataset 4.2), suggests that glycolysis was enhanced in *T. pseudonana*. Starch stores were depleted, as indicated by significantly lowered concentration of maltopentose, amylopectin, cellopentose, and/or verbascose (-3.2-fold) (Table 4.2, Table 4.5), whereas concentrations of small (e.g., C_8-C_{10}) fatty acids increased (Table 4.4). Several fatty acid biosynthesis and metabolism enzymes (Table 4.1) were more abundant in *T. pseudonana* in response to *K. brevis* allelopathy, including

acetyl CoA carboxylase (+2.3 fold) and acyl CoA dehydrogenase (+3.0 fold) (Dataset 4.2). Together, this suggests that allelopathy caused enhanced glycolysis in *T. pseudonana*, while also stimulating synthesis and β-oxidation of fatty acids. Allelopathy may have altered the photosynthetic capability of *T. pseudonana*, as evidenced by the decreased abundance of 12 photosynthetic proteins, even though the abundance of RuBisCO, the key enzyme responsible for carbon fixation, remained unchanged (Table 4.1, Dataset 4.2). The reduction in photosystem complexes and surge in fatty acid β-oxidation may indicate that cells experiencing allelopathy became carbon-limited.

Table 4.4: Candidate metabolites identified from ^{1}H and HSQC NMR spectra of *Thalassiosira pseudonana* exposed to *Karenia brevis* allelopathy. Chemical shifts in bold indicate observed signals in *T. pseudonana* extracts, while non-bold shifts indicate signals not observed due to overlap in that spectral region. Chemical shift values that were not observed are reported from either Boroujerdi et al.[45] or the human metabolome database (HMDB). Metabolic pathways reported from KEGG.

Metabolite	^{1}H chemical shift (ppm)	^{13}C chemical shift (ppm)	Effect of *K. brevis*	Metabolic Pathway/Function
Acetate	**1.93**	26.2	enhanced	central energy metabolism
Dimethylamine[1]	**2.71**	37.3	enhanced	nitrogen metabolism
Fatty acid(s)	**0.85, 1.27, 1.53, 2.16**	39.5, 27.9, 30.9, 23.9, 33.2, 14.6	enhanced	fatty acid metabolism
Sarcosine[1]	**2.72**, 3.59	35.6, 53.5	enhanced	nitrogen metabolism
Glutamate	**2.07**, 2.14, **2.35**, **3.78**	**31.7**, 29.9, **36.1**, **57.4**	enhanced	amino acid metabolism, energy metabolism
Proline	**2.05**, **2.07**, **2.36**, **3.35**, **3.43**, **4.15**	**26.5**, **31.7**, **31.6**, **48.8**, **49.1**, **63.9**	enhanced	amino acid metabolism, urea cycle, osmoprotectant
Taurine	**3.29**, 3.43	**51.7**, 38.3	suppressed	
Alanine	1.45, **3.80**	19.1, **53.6**	suppressed	
Betaine	**3.28**, **3.91**	**56.2**, **68.8**	suppressed	osmoprotectant
Dihydrouracil	**2.74**, **3.48**	**34.0**, **43.1**	suppressed	pyrimidine catabolism, nitrogen recycling
Glucose[2]	**3.33**, **3.36**, **3.53**, **3.59**, **3.57**, 3.76, **3.79**, **3.87**, **3.92**, 3.93, 4.66, 5.25	**79.6**, **76.3**, **78.3**, **75.9**, **75.0**, 74.0, **63.5**, **74.0**, **63.1**, 73.6, 98.9, 95.1	suppressed	central energy metabolism
Myo-inositol[2]	**3.33**, **3.59**, **3.57**, 3.75, 3.75, **4.08**	**79.6**, **75.9**, **74.0**, 75.7, 73.9, **74.5**	suppressed	central energy metabolism
Glycerate	**3.70, 3.79, 4.17**	**67.7, 67.7, 77.8**	suppressed	photorespiration
Homarine	**4.37**, 7, 8.04, **8.55**, **8.73**	**49.3**, 130.2, 129.0, **149.5**, **148.6**	suppressed	osmoprotectant

[1] indicates metabolites share observed chemical shifts
[2] indicates metabolites share multiple observed chemical shifts

Table 4.5: Candidate metabolites tentatively identified from UPLC-MS data of *Thalassiosira pseudonana* exposed to *Karenia brevis* allelopathy.

Retention Time (min)	m/z	Metabolite Candidates in Database	Molecular Formula	Mass Error (mDa)	Detected Ion Type	Effect of *K. brevis*	Metabolic Pathway/Function
4.21	557.2721	*Lipid*	$C_{31}H_{42}O_9$	1.1	$[M-H]^-$	Enhanced	Membrane constituent [46]
		Phospholipid	$C_{24}H_{47}O_{12}P$	3.5			Membrane lipid [47]
4.54	581.2994	Unknown	$C_{31}H_{42}N_4O_7$	1.3	$[M-H]^-$	Enhanced	-
4.30	529.2555	PG(20:5/0:0)	$C_{26}H_{43}O_9P$	1.7	$[M-H]^-$	Suppressed	Membrane lipid [47]
		Lipid	$C_{26}H_{42}O_{11}$	9.9			Membrane constituent [46]
2.68	701.1880	*Polyphenol*	$C_{31}H_{38}O_{16}$	2.6	$[M+Cl]^-$	Suppressed	Oxidative stress protectant [48]
		Mannan	$C_{24}H_{42}O_{21}$	3.3			Cell wall component [49]
		Stachyose	$C_{24}H_{42}O_{21}$				Galactose metabolism[1]
		Cellotetraose	$C_{24}H_{42}O_{21}$				Starch & sucrose metabolism[1]
5.01	735.3346	*Alkaloid*	$C_{37}H_{52}N_2O_{11}$	8.1	$[M+Cl]^-$	Suppressed	Nitrogen storage [50]; allelochemical [51]
2.77[a]	827.2672	Maltopentaose	$C_{30}H_{52}O_{26}$	0.2	$[M-H]^-$	Suppressed	Glucose storage [52]

Table 4.5 (continued).

		Amylopectin	$C_{30}H_{52}O_{26}$				Glucose/starch storage [53]
		Cellopentaose	$C_{30}H_{52}O_{26}$				Starch & sucrose metabolism[1]
		Verbascose	$C_{30}H_{52}O_{26}$				Carbohydrate storage [54]
4.19	1113.4600	*Terpene glycoside*	*$C_{50}H_{78}O_{25}$*	7.4	$[M+Cl]^-$	Suppressed	Membrane constituent [46]

[1]biological function information obtained from KEGG
[a]the chlorine adduct of this metabolic feature was also selected as a discriminant feature
italics indicate compounds not previously described from marine planktonic organism

4.6.1 Cell Protection Pathways Impacted

In previous work, *K. brevis* exudates were found to increase cell membrane permeability of several competitors, via unknown mechanisms.[17] In the current study, metabolic and biosynthetic activities associated with cell membranes were altered in *T. pseudonana* in response to *K. brevis* allelopathy, as evidenced by decreased mannan and C_{26} phospholipid or C_{31} lipid concentrations, as well as an increase in a C_{24} phospholipid or C_{26} lipid (Table 4.2, Table 4.5). This is suggestive of membrane restructuring and a possible increase in permeability since mannans and phospholipids are primary membrane components of *T. pseudonana*[49] and other photosynthetic organisms.[55] These changes could be the result of membrane instability and the subsequent attempt of *T. pseudonana* to alter membrane phospholipid content via increased biosynthesis. Recently, Martin et al. revealed that phosphorus-deprived *T. pseudonana* substitute betaine lipids and sulfolipids for phospholipids in cellular membranes.[47] In addition, Riekhof et al. demonstrated that betaine lipid synthesis is associated with the degradation of membrane phospholipids.[56] These coordinated pathways provide a possible explanation to the significantly reduced betaine concentrations (-0.58-fold) since betaine would be consumed to build betaine-based lipids (Table 2, Table 4.4). Enzymes involved in sulfolipid biosynthesis were present in higher concentrations in *T. pseudonana* responding to *K. brevis* (e.g., sulfolipid biosynthesis protein; Dataset 4.2), suggesting a modification of *T. pseudonana*'s content via an increase in sulfolipids. Although phosphate was not limiting to *T. pseudonana* in this experiment, it is possible that *K. brevis* allelopathy reduced the capacity of *T. pseudonana* to utilize nutrients present in the media. Plankton lipids mediate a number of important ecological interactions, including

grazing,[57] susceptibility to viral infection,[58] and toxin sensitivity.[59] Thus, future experiments testing how phytoplankton competition alters lipid biochemistry and membrane stability could lead to discovery of cascading ecological effects on other types of interactions.

Another critical aspect of membrane structure and function is the ability of a cell to control solute osmosis. From the current study it is apparent that osmotic regulation was impacted by *K. brevis* allelopathy: concentrations of homarine (-2.5-fold) and betaine were suppressed when *T. pseudonana* was exposed to *K. brevis* (Table 4.2, Table 4.4) and are recognized osmolytes in marine planktonic organisms.[60, 61] *T. pseudonana* may respond to *K. brevis* allelopathy by compensating for impeded osmotic regulation by enhancing or maintaining concentrations of both proline and dimethylamine (Table 4.2).[21, 62]

When *T. pseudonana* was exposed to *K. brevis* allelopathy, multiple enzymes involved in critical pathways related to oxidative stress were enhanced. In addition, several enzymes responsible for mitigating high oxidative stress were more abundant. Fifteen oxidoreductases increased in abundance, including manganese superoxide dismutase and ascorbate peroxidase (Table 4.1, Dataset 4.2). Our oPLS-DA model suggests glycerate concentrations were reduced, as part of a group of compounds whose concentrations covaried, in *T. pseudonana* exposed to *K. brevis* indicating that photorespiration was disrupted (Table 4.2), which would also likely increase oxidative stress.[63, 64] Recently, Schlegel et al. proposed that an increase in oxidative stress may force cells to reduce the abundance of photosynthetic and electron carrier enzymes.[48] Consistent with that prediction, in the current study proteins involved in electron carrier

activity were suppressed, whereas proton transport activity was enhanced, suggesting potential compensation within *T. pseudonana* cells (Table 4.1, Dataset 4.2). Seven proteins involved in the pentose phosphate pathway, a side-step to glycolysis that yields reducing equivalents of NADPH,[19] yielded heightened abundances (Table 4.1, Dataset 4.2). Recent findings also show that diatoms experiencing oxidative stress lose their ability to assimilate nitrogen, due to either altered redox states of critical nitrogen assimilation enzymes, or the loss of reducing equivalents (e.g. NADPH).[65] This suggests that in the current study, oxidatively stressed *T. pseudonana* may have been forced to recycle internal nitrogen stores as a downstream effect of exposure to *K. brevis* allelopathy. We also observed alterations in the concentrations of several amino acids (increased concentrations of glutamate and proline; decreased concentrations of taurine; Table 4.2, Table 4.4) and an unidentified alkaloid (Table 4.2, Table 4.5), suggesting a reshuffling and possible recycling of internal nitrogen stores. Together, this suggests that *T. pseudonana* was responding to increased oxidative stress as a result of *K. brevis* allelopathy.

4.6.2 Cells in Crisis

Overall, it appears that *T. pseudonana* cells exposed to *K. brevis* allelopathy experienced substantially disrupted metabolic processes indicative of heightened stress (Figure 4.4). In addition to the metabolic and proteomic changes described above, we observed a fourfold increase in the abundance of the diatom's trypsin with a concomitant increase in the abundance of an anti-trypsin protease inhibitor as well as reduced concentrations of a serine protease inhibitor (Dataset 4.2). For trypsin annotation, a total of 19 unique peptides were identified, none of which resembled the pure porcine trypsin

used for digestion during sample preparation. These findings suggest that protein degradation pathways were induced which may approach conditions typically observed for programmed cell death response in diatoms.[66] In addition, concentrations of several DNA replication enzymes, as well as eight chromatin assembly proteins were suppressed (e.g., histone-3: -3.4 fold; Table 4.1, Dataset 4.2), consistent with a strong negative impact of allelopathy on *T. pseudonana* growth (Figure 4.1).

Mucins, gel-like glycoproteins, were in higher abundance in *T. pseudonana* exposed to allelopathy (Dataset 4.2). While largely unexplored for phytoplankton, in other organisms mucins have been found to reduce adherence of bacteria.[67] If diatom mucins prevent colonization by harmful bacteria or aggregation of diatom cells, this may represent a valuable defense mechanism or a response to stress. Furthermore, if mucins are shown to be involved in diatom aggregation, then allelopathy may directly influence the flux of particulate carbon from the water column to depth in the ocean. Mucins have also been found to demonstrate antiapoptopic activity.[68, 69] Previous studies on *T. pseudonana* have found ascorbate peroxidase genes to be up-regulated in iron-limited situations, whereas mucin genes were down-regulated, demonstrating a balance between ROS reducing mechanisms and the initiation of programmed cell death pathways in stressful situations.[70] This balancing act ultimately allows acclimation to stress, which may be disrupted by allelopathy.

4.7 Conclusion

Whole cell metabolomic and proteomic analyses revealed differing responses of planktonic competitors to *K. brevis* allelopathy, suggesting that co-occurring species may have evolved partial resistance to allelopathy via robust metabolic pathways. In contrast,

a “naïve” competitor, *T. pseudonana*, which does not co-occur with *K. brevis* blooms, suffered greater metabolic disruption and growth suppression when exposed to *K. brevis* allelopathy. Critical metabolic processes including glycolysis, photosynthesis, cell membrane maintenance, osmoregulation, and responses to oxidative stress were all impacted in this sensitive competitor. If other sensitive competitors are also found to undergo similarly altered metabolic pathways, then allelopathy among phytoplankton may affect ecosystem level processes even more than previously believed due to the importance of phytoplankton on carbon fixation and nutrient dynamics in the world’s oceans. This interdisciplinary systems biology approach provides an unbiased opportunity to establish novel, testable hypotheses towards understanding the mechanisms by which allelopathy alters species composition in plankton communities, and, thus, the roles of chemical cues in mediating important ecological interactions.

4.8 References

(1) Field, C. B.; Behrenfeld, M. J.; Randerson, J. T.; Falkowski, P., Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* **1998,** *281* (5374), 237-240.

(2) Strom, S. L., Microbial ecology of ocean biogeochemistry: a community perspective. *Science* **2008,** *320* (5879), 1043-1045.

(3) Kubanek, J.; Hicks, M. K.; Naar, J.; Villareal, T. A., Does the red tide dinoflagellate *Karenia brevis* use allelopathy to outcompete other phytoplankton? *Limnol. Oceanogr.* **2005,** *50* (3), 883-895.

(4) Tillmann, U.; John, U.; Cembella, A., On the allelochemical potency of the marine dinoflagellate *Alexandrium ostenfeldii* against heterotrophic and autotrophic protists. *J. Plankton Res.* **2007,** *29* (6), 527-543.

(5) Sheng, J.; Malkiel, E.; Katz, J.; Adolf, J. E.; Place, A. R., A dinoflagellate exploits toxins to immobilize prey prior to ingestion. *Proc. Natl. Acad. Sci. U. S. A.* **2010,** *107* (5), 2082-2087.

(6) Inderjit; Wardle, D. A.; Karban, R.; Callaway, R. M., The ecosystem and evolutionary contexts of allelopathy. *Trends Ecol. Evol.* **2011,** *26* (12), 655-662.

(7) Thacker, R. W.; Becerro, M. A.; Lumbang, W. A.; Paul, V. J., Allelopathic interactions between sponges on a tropical reef. *Ecology* **1998,** *79* (5), 1740-1750.

(8) Rasher, D. B.; Hay, M. E., Chemically rich seaweeds poison corals when not controlled by herbivores. *Proc. Natl. Acad. Sci. U. S. A.* **2010,** *107* (21), 9683-9688.

(9) Ma, H. Y.; Krock, B.; Tillmann, U.; Cembella, A., Preliminary characterization of extracellular allelochemicals of the toxic marine dinoflagellate *Alexandrium tamarense* using a *Rhodomonas salina* bioassay. *Mar. Drugs* **2009,** *7* (4), 497-522.

(10) Vardi, A.; Schatz, D.; Beeri, K.; Motro, U.; Sukenik, A.; Levine, A.; Kaplan, A., Dinoflagellate-cyanobacterium communication may determine the composition of phytoplankton assemblage in a mesotrophic lake. *Curr. Biol.* **2002,** *12* (20), 1767-1772.

(11) Suikkanen, S.; Engstrom-Ost, J.; Jokela, J.; Sivonen, K.; Viitasalo, M., Allelopathy of Baltic Sea cyanobacteria: no evidence for the role of nodularin. *J. Plankton Res.* **2006,** *28* (6), 543-550.

(12) Ribalet, F.; Berges, J. A.; Ianora, A.; Casotti, R., Growth inhibition of cultured marine phytoplankton by toxic algal-derived polyunsaturated aldehydes. *Aquat. Toxicol.* **2007,** *85* (3), 219-227.

(13) Tillmann, U.; John, U., Toxic effects of *Alexandrium spp.* on heterotrophic dinoflagellates: an allelochemical defence mechanism independent of PSP-toxin content. *Mar. Ecol. Prog. Ser.* **2002,** *230*, 47-58.

(14) Landsberg, J. H.; Flewelling, L. J.; Naar, J., *Karenia brevis* red tides, brevetoxins in the food web, and impacts on natural resources: Decadal advancements. *Harmful Algae* **2009,** *8* (4), 598-607.

(15) Poulson, K. L.; Sieg, R. D.; Prince, E. K.; Kubanek, J., Allelopathic compounds of a red tide dinoflagellate have species-specific and context-dependent impacts on phytoplankton. *Mar. Ecol. Prog. Ser.* **2010,** *416*, 69-78.

(16) Prince, E. K.; Poulson, K. L.; Myers, T. L.; Sieg, R. D.; Kubanek, J., Characterization of allelopathic compounds from the red tide dinoflagellate *Karenia brevis*. *Harmful Algae* **2010,** *10* (1), 39-48.

(17) Prince, E. K.; Myers, T. L.; Kubanek, J., Effects of harmful algal blooms on competitors: allelopathic mechanisms of the red tide dinoflagellate *Karenia brevis*. *Limnol. Oceanogr.* **2008,** *53* (2), 531-541.

(18) Schmid, A.; Blank, L. M., Systems biology: hypothesis-driven omics integration. *Nat. Chem. Biol.* **2010,** *6* (7), 485-487.

(19) Nunn, B. L.; Faux, J. F.; Hippmann, A. A.; Maldonado, M. T.; Harvey, H. R.; Goodlett, D. R.; Boyd, P. W.; Strzepek, R. F., Diatom proteomics reveals unique acclimation strategies to mitigate Fe limitation. *PLoS One* **2013,** *8* (10), e75653.

(20) Carvalho, R. N.; Lettieri, T., Proteomic analysis of the marine diatom *Thalassiosira pseudonana* upon exposure to benzo(a)pyrene. *BMC Genomics* **2011,** *12*, 159.

(21) Bromke, M. A.; Giavalisco, P.; Willmitzer, L.; Hesse, H., Metabolic analysis of adaptation to short-term changes in culture conditions of the marine diatom *Thalassiosira pseudonana*. *PLoS One* **2013,** *8* (6), e67340.

(22) Prince, E. K.; Irmer, F.; Pohnert, G., Domoic acid improves the competitive ability of *Pseudo-nitzschia delicatissima* against the diatom *Skeletonema marinoi*. *Mar. Drugs* **2013,** *11* (7), 2398-412.

(23) Poulson-Ellestad, K. L.; Jones, C. M.; Roy, J.; Viant, M. R.; Fernández, F. M.; Kubanek, J.; Nunn, B. L., Metabolomics and proteomics reveal impacts of chemically

mediated competition on marine plankton. *Proc. Natl. Acad. Sci. U. S. A.* **2014,** *111* (24), 9009-9014.

(24) Rajalahti, T.; Arneberg, R.; Berven, F. S.; Myhr, K.-M.; Ulvik, R. J.; Kvalheim, O. M., Biomarker discovery in mass spectral profiles by means of selectivity ratio plot. *Chemometrics Intellig. Lab. Syst.* **2009,** *95* (1), 35-48.

(25) Kanehisa, M.; Goto, S., KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **2000,** *28* (1), 27-30.

(26) Smith, C. A.; O'Maille, G.; Want, E. J.; Qin, C.; Trauger, S. A.; Brandon, T. R.; Custodio, D. E.; Abagyan, R.; Siuzdak, G., METLIN: a metabolite mass spectral database. *Ther. Drug Monit.* **2005,** *27* (6), 747-751.

(27) Sud, M.; Fahy, E.; Cotter, D.; Brown, A.; Dennis, E. A.; Glass, C. K.; Merrill, A. H.; Murphy, R. C.; Raetz, C. R. H.; Russell, D. W.; Subramaniam, S., LMSD: LIPID MAPS structure database. *Nucleic Acids Res.* **2007,** *35* (suppl 1), D527-D532.

(28) Cui, Q.; Lewis, I. A.; Hegeman, A. D.; Anderson, M. E.; Li, J.; Schulte, C. F.; Westler, W. M.; Eghbalnia, H. R.; Sussman, M. R.; Markley, J. L., Metabolite identification via the Madison metabolomics consortium database. *Nat. Biotechnol.* **2008,** *26* (2), 162-164.

(29) Wishart, D. S.; Tzur, D.; Knox, C.; Eisner, R.; Guo, A. C.; Young, N.; Cheng, D.; Jewell, K.; Arndt, D.; Sawhney, S., HMDB: the human metabolome database. *Nucleic Acids Res.* **2007,** *35* (suppl 1), D521-D526.

(30) Zhou, B.; Wang, J.; Ressom, H. W., MetaboSearch: tool for mass-based metabolite identification using multiple databases. *PLoS One* **2012,** *7* (6), e40096.

(31) Douglas, D. J.; Konenkov, N. V., Ion source emittance influence on the transmission of a quadrupole operated in the second stability region. *J. Am. Soc. Mass. Spectrom.* **1998,** *9* (10), 1074-1080.

(32) Sasso, S.; Pohnert, G.; Lohr, M.; Mittag, M.; Hertweck, C., Microalgae in the postgenomic era: a blooming reservoir for new natural products. *FEMS Microbiol. Rev.* **2012,** *36* (4), 761-785.

(33) Chen, E. I.; Cociorva, D.; Norris, J. L.; Yates, J. R., Optimization of mass spectrometry-compatible surfactants for shotgun proteomics. *J. Proteome Res.* **2007,** *6* (7), 2529-2538.

(34) Choi, H.; Fermin, D.; Nesvizhskii, A. I., Significance analysis of spectral count data in label-free shotgun proteomics. *Mol. Cell. Proteomics* **2008,** *7* (12), 2373-2385.

(35) Eng, J. K.; Jahan, T. A.; Hoopmann, M. R., Comet: an open-source MS/MS sequence database search tool. *Proteomics* **2013,** *13* (1), 22-24.

(36) Eng, J. K.; Fischer, B.; Grossmann, J.; MacCoss, M. J., A Fast SEQUEST Cross Correlation Algorithm. *J. Proteome Res.* **2008,** *7* (10), 4598-4602.

(37) Elias, J. E.; Gygi, S. P., Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods* **2007,** *4* (3), 207-214.

(38) Holm, L.; Sander, C., Removing near-neighbour redundancy from large protein sequence collections. *Bioinformatics* **1998,** *14* (5), 423-429.

(39) Ison, J. EMBOSS: skipredundant. http://emboss.sourceforge.net/apps/release/6.1/emboss/apps/skipredundant.html

(40) Sherman, B. T.; Lempicki, R. A., Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **2009,** *37* (1), 1-13.

(41) Da Wei Huang, B. T. S.; Lempicki, R. A., Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **2008,** *4* (1), 44-57.

(42) Robertson, D. G.; Reily, M. D.; Baker, J. D., Metabonomics in pharmaceutical discovery and development. *J. Proteome Res.* **2007,** *6* (2), 526-539.

(43) Zdzislaw, A. W., Biochemistry of Phytosterol Conjugates. In *Physiology and Biochemistry of Sterols*, AOCS Publishing: 1992; pp 361-395.

(44) Badylak, S.; Phlips, E. J.; Baker, P.; Fajans, J.; Boler, R., Distributions of phytoplankton in Tampa Bay Estuary, USA 2002-2003. *Bull. Mar. Sci.* **2007,** *80* (2), 295-317.

(45) Boroujerdi, A. F. B.; Lee, P. A.; DiTullio, G. R.; Janech, M. G.; Vied, S. B.; Bearden, D. W., Identification of isethionic acid and other small molecule metabolites of *Fragilariopsis cylindrus* with nuclear magnetic resonance. *Anal. Bioanal. Chem.* **2012,** *404* (3), 777-784.

(46) Pozzi, G.; Birault, V.; Werner, B.; Dannenmuller, O.; Nakatani, Y.; Ourisson, G.; Terakawa, S., Single-chain polyprenyl phosphates form “primitive” membranes. *Angew. Chem. Int. Ed.* **1996,** *35* (2), 177-180.

(47) Martin, P.; Van Mooy, B. A.; Heithoff, A.; Dyhrman, S. T., Phosphorus supply drives rapid turnover of membrane phospholipids in the diatom *Thalassiosira pseudonana*. *ISME J.* **2010,** *5* (6), 1057-1060.

(48) Schlegel, K.; Welte, C.; Deppenmeier, U.; Müller, V., Electron transport during aceticlastic methanogenesis by *Methanosarcina acetivorans* involves a sodium-translocating Rnf complex. *FEBS J.* **2012,** *279* (24), 4444-4452.

(49) Chiovitti, A.; Harper, R. E.; Willis, A.; Bacic, A.; Mulvaney, P.; Wetherbee, R., Variations in the substituted 3-linked mannans closely associated with the silicified walls of diatoms. *J. Phycol.* **2005,** *41* (6), 1154-1161.

(50) Wink, M.; Witte, L., Quinolizidine alkaloids as nitrogen source for lupin seedlings and cell cultures. *Z. Naturforsch* **1985**, 767-775.

(51) Wink, M., Evolution of secondary metabolites from an ecological and molecular phylogenetic perspective. *Phytochemistry* **2003,** *64* (1), 3-19.

(52) Critchley, J. H.; Zeeman, S. C.; Takaha, T.; Smith, A. M.; Smith, S. M., A critical role for disproportionating enzyme in starch breakdown is revealed by a knock-out mutation in Arabidopsis. *Plant J.* **2001,** *26* (1), 89-100.

(53) Nag, A.; Lunacek, M.; Graf, P. A.; Chang, C. H., Kinetic modeling and exploratory numerical simulation of chloroplastic starch degradation. *BMC Syst. Biol.* **2011,** *5* (1), 94.

(54) Avigad, G.; Dey, P., Carbohydrate metabolism: storage carbohydrates. In *Plant Biochemistry*, Academic Press: 1997; pp 143-204.

(55) Popper, Z. A.; Michel, G.; Hervé, C.; Domozych, D. S.; Willats, W. G. T.; Tuohy, M. G.; Kloareg, B.; Stengel, D. B., Evolution and diversity of plant cell walls: from algae to flowering plants. *Annu. Rev. Plant Biol.* **2011,** *62* (1), 567-590.

(56) Riekhof, W. R.; Andre, C.; Benning, C., Two enzymes, BtaA and BtaB, are sufficient for betaine lipid biosynthesis in bacteria. *Arch. Biochem. Biophys.* **2005,** *441* (1), 96-105.

(57) Miralto, A.; Barone, G.; Romano, G.; Poulet, S. A.; Ianora, A.; Russo, G. L.; Buttino, I.; Mazzarella, G.; Laabir, M.; Cabrini, M.; Giacobbe, M. G., The insidious effect of diatoms on copepod reproduction. *Nature* **1999,** *402* (6758), 173-176.

(58) Vardi, A.; Van Mooy, B. A. S.; Fredricks, H. F.; Popendorf, K. J.; Ossolinski, J. E.; Haramaty, L.; Bidle, K. D., Viral glycosphingolipids induce lytic infection and cell death in marine phytoplankton. *Science* **2009,** *326* (5954), 861-865.

(59) Deeds, J. R.; Place, A. R., Sterol-specific membrane interactions with the toxins from *Karlodinium micrum* (Dinophyceae) — a strategy for self-protection? *Afr. J. Mar. Sci.* **2006,** *28* (2), 421-425.

(60) Dickson, D. M. J.; Kirst, G. O., The role of beta-dimethylsulphoniopropionate, glycine betaine and homarine in the osmoacclimation of *Platymonas-subcordiformis*. *Planta* **1986,** *167* (4), 536-543.

(61) Keller, M. D.; Kiene, R. P.; Matrai, P. A.; Bellows, W. K., Production of glycine betaine and dimethylsulfoniopropionate in marine phytoplankton. I. Batch cultures. *Mar. Biol.* **1999,** *135* (2), 237-248.

(62) Dickson, D. M. J.; Kirst, G. O., Osmotic adjustment in marine eukaryotic algae: the role of inorganic ions, quaternary ammonium, tertiary sulphonium and carbohydrate solutes. *New Phytol.* **1987,** *106* (4), 645-655.

(63) Allen, A. E.; LaRoche, J.; Maheswari, U.; Lommer, M.; Schauer, N.; Lopez, P. J.; Finazzi, G.; Fernie, A. R.; Bowler, C., Whole-cell response of the pennate diatom *Phaeodactylum tricornutum* to iron starvation. *Proc. Natl. Acad. Sci. U. S. A.* **2008,** *105* (30), 10438-10443.

(64) Kroth, P. G.; Chiovitti, A.; Gruber, A.; Martin-Jezequel, V.; Mock, T.; Parker, M. S.; Stanley, M. S.; Kaplan, A.; Caron, L.; Weber, T.; Maheswari, U.; Armbrust, E. V.; Bowler, C., A model for carbohydrate metabolism in the diatom *Phaeodactylum tricornutum* deduced from comparative whole genome analysis. *PLoS One* **2008,** *3* (1), e1426.

(65) Rosenwasser, S.; van Creveld, S. G.; Schatz, D.; Malitsky, S.; Tzfadia, O.; Aharoni, A.; Levin, Y.; Gabashvili, A.; Feldmesser, E.; Vardi, A., Mapping the diatom redox-sensitive proteome provides insight into response to nitrogen stress in the marine environment. *Proc. Natl. Acad. Sci. U. S. A.* **2014,** *111* (7), 2740-2745.

(66) Bidle, K. D.; Bender, S. J., Iron starvation and culture age activate metacaspases and programmed cell death in the marine diatom *Thalassiosira pseudonana*. *Eukaryot. Cell* **2008,** *7* (2), 223-236.

(67) Caldara, M.; Friedlander, R. S.; Kavanaugh, N. L.; Aizenberg, J.; Foster, K. R.; Ribbeck, K., Mucin biopolymers prevent bacterial aggregation by retaining cells in the free-swimming state. *Curr. Biol.* **2012,** *22* (24), 2325-30.

(68) Singh, P. K.; Hollingsworth, M. A., Cell surface-associated mucins in signal transduction. *Trends Cell Biol.* **2006,** *16* (9), 467-476.

(69) Workman, H. C.; Sweeney, C.; Carraway, K. L., III, The membrane mucin muc4 inhibits apoptosis induced by multiple insults via erbb2-dependent and erbb2-independent mechanisms. *Cancer Res.* **2009,** *69* (7), 2845-2852.

(70) Thamatrakoln, K.; Korenovska, O.; Niheu, A. K.; Bidle, K. D., Whole-genome expression analysis reveals a role for death-related genes in stress acclimation of the diatom *Thalassiosira pseudonana*. *Environ. Microbiol.* **2012,** *14* (1), 67-81.

CHAPTER 5. TRANSMISSION MODE DIRECT ANALYSIS IN REAL TIME MASS SPECTROMETRY FOR FAST UNTARGETED METABOLIC FINGERPRINTING

Adapted with permission from
Jones, C. M.; Fernández, F. M., Transmission Mode Direct Analysis in Real Time Mass Spectrometry for Fast Untargeted Metabolic Fingerprinting. *Rapid Commun. Mass Spectrom.* **2013**, *27* (12), 1311-1318. Copyright © 2013 John Wiley & Sons, Ltd.

Jones, C. M., Monge, M. E., Fernández, F. M., *Metabolite Profiling by Direct Analysis in Real Time Mass Spectrometry* In Mass Spectrometry Methods in Metabolomics, Methods in Molecular Biology. Humana Press USA, Springer Science+Business Media: New York, 2014. pp. 275-289. Copyright © 2014 Springer Science+Business Media.

5.1 Abstract

Although current hyphenated mass spectrometry (MS) techniques are very powerful for metabolomics studies, there is a clear need for more rapid, high-throughput MS approaches. A rapid metabolic fingerprinting method that utilizes a new transmission mode direct analysis in real time (TM-DART) sampling technique coupled with quadrupole time of flight mass spectrometry (QTOF MS) is presented. In this approach, the sample is deposited directly on a stainless steel mesh that is held in the ionization region by a custom-built module. As a result, the DART plasma gas stream interacts with the sample in a flow-through fashion, which maximizes the sample-ionizing species interaction and minimizes variance in sample positioning. The optimization of TM-DART parameters directly affecting metabolite desorption and ionization, such as sample position and ionizing gas desorption temperature, was critical in achieving high sensitivity and detecting a broad mass range of metabolites. Ramping the ionizing gas desorption temperature further enhanced analysis by adding a simple separation

dimension to this ambient approach. In terms of reproducibility, TM-DART compared favorably with traditional probe mode (PM-) DART analysis, with coefficients of variation as low as 16%. The longer-lasting TM-DART signals enabled the acquisition of full scan and product ion accurate mass spectra in a single experiment, resulting in greater confidence in metabolite identification. TM-DART MS proved to be a powerful analytical technique for rapid metabolome analysis of human blood sera and was adapted for exhaled breath condensate (EBC) analysis. To determine the feasibility of utilizing TM-DART for metabolomics investigations, TM-DART was interfaced with traveling wave ion mobility spectrometry (TWIMS) time-of-flight (TOF) MS for the analysis of EBC samples from cystic fibrosis (CF) patients and healthy controls. TM-DART-TWIMS-TOF MS was able to successfully detect CF in this small sample cohort.

5.2 Metabolic Fingerprinting by Direct Analysis in Real Time Mass Spectrometry

5.2.1 Direct Analysis in Real Time

Direct Analysis in Real Time (DART)[1] belongs to the growing family of direct sampling "ambient" ionization techniques. Ambient MS techniques make use of well-established ionization principles, such as electrospray ionization (ESI), chemical ionization (CI), and photo-ionization (PI), in an open air format which allows high throughput experiments to be performed with minimal sample preparation. Non-destructive surface analysis, spatially-resolved analyte detection, multi-analyte fingerprinting, matrix-less molecular imaging, and selective ionization enhancement by

means of specific gas-phase chemical reactions are possible with ambient MS approaches.[2]

DART uses a DC point-to-plane atmospheric pressure electric discharge in the glow regime to generate metastable species from a flowing gas in a chamber that is physically separated from the analyte ionization region. The gas stream with metastables is heated and directed through a set of one or more grid electrodes that prevent most plasma ionic species from interacting directly with the sample, and minimize ion-ion and ion-electron recombination of ionic species generated within the external ionization region. The metastable species formed within the discharge supporting gas, generally He or N_2, generate protonated water clusters by Penning ionization of atmospheric gases, which act as reactant ions that chemically ionize thermally-desorbed analyte molecules through proton-transfer reactions.[3] The thermal conductivity of the gas used for desorption is a critical parameter affecting the sensitivity due to the thermal desorption step, with He being approximately one order of magnitude more conducting than N_2. In the majority of applications, the utilization of DART has mainly been focused on analytes below 1 kDa, which is well suited for most metabolomic experiments.[4-8]

In the traditional probe mode (PM)-DART sample introduction method, liquid samples are deposited on a melting point capillary, allowed to dry, and then introduced into the ionization region mechanically or manually. In this approach, a multitude of geometric variables affect analyte desorption, ion transmission, and the reproducibility of the experiments, as the sample capillary itself creates turbulence while being introduced into the ionization region.[9] To minimize the variance introduced by the various DART geometric factors, our group was first to propose a transmission mode (TM)-DART

geometry [10] as an alternative to the PM method. TM operation is highly advantageous as the sample can be deposited directly on a stainless steel mesh that is held in the ionization region by a custom-built module placed directly in between the ion source and the mass spectrometer inlet. TM operation enables the DART plasma gas stream to interact with the sample in a flow-through fashion, therefore maximizing interaction with ionizing species, and minimizing variance in sample positioning.

5.2.2 Direct Analysis in Real Time in Metabolomics

DART applications in the field of metabolomics have shown a steady growth since the first reports in 2010,[4, 11] and are expected to become more prominent. Compared to analytical tools used routinely in metabolomics such as continuous infusion electrospray ionization (ESI)-MS, capillary electrophoresis (CE)-MS, liquid chromatography (LC)-MS and gas chromatography (GC)-MS,[12-14] DART-MS is attractive due to several reasons: (1) sample analysis can be completed in a few minutes without chromatographic separation, (2) memory effects are minimized because all parts in contact with the sample are disposable, (3) analytes with a broad spectrum of polarities are ionized because of the acidity of the chemical ionization reagent ions involved,[15] and (4) simpler spectra than ESI are produced, as a minimal number of adducts/multimers is generated.[7, 11] However, the lack of chromatographic separation increases the probability of overlapping chemical features and ionization suppression, making the use of high mass resolution instrumentation a necessity to achieve the required peak capacity and find elemental formulae for unknown metabolites with a higher degree of confidence.[16]

5.2.3 Ion Mobility Spectrometry

Ion mobility spectrometry (IMS) techniques are appealing for their ability to separate ionized analytes, even isobars, on a millisecond timescale based on differences in their gas-phase mobility within an electric field, which is affected by mass, charge, field strength, and collisional cross section.[17] Each analyte has a characteristic drift velocity (v_d) that is proportional to the applied electric field strength (E) and ion mobility constant (K):

$$v_d = KE \qquad [1]$$

Ion mobility values are typically normalized to standard temperature and pressure conditions and reported as reduced mobility (K_0). The time required for analytes to reach the detector (drift time, t_d) can be used to determine ion mobility values:

$$K = \frac{L^2}{t_d U} \qquad [2]$$

where L is the drift length and U is the potential difference of the electric field. The Mason-Schamp equation[18] can be used to calculate collision cross sections (Ω) based on drift times:

$$\Omega = \frac{3ze}{16N}\left(\frac{2\pi}{\mu k_B T}\right)^{1/2}\frac{1}{K_0} \qquad [3]$$

where z is the charge state of the ion, e is the elementary charge, N is the drift gas number density, μ is the reduced mass of the ion-neutral drift gas pair, k_B is the Boltzmann constant, and T is the gas temperature. Collision cross sections, which are indicative of size and shape, provide characteristic structural information for analytes.

IMS coupled to MS provides a separation dimension for the analysis of complex biological samples. IM-MS has been successfully employed for metabolomics studies.[19-

[21] While not completely orthogonal to MS, IMS has been shown to increase the peak capacity of MS sixfold.[20] The coupling of DART to high-resolution ion mobility TOF MS may also alleviate the lack of chromatographic separation in DART metabolomics studies. Various IMS techniques are currently available, including drift tube IMS,[22] differential mobility spectrometry (DMS),[23] and traveling wave ion mobility spectrometry (TWIMS),[24] each offering varying degrees of separation power and signal-to-noise ratio enhancement. A traveling voltage wave is applied to a stacked ring ion guide in the TWIMS technique (Figure 5.1). Adjacent electrodes have radially confining radiofrequency voltages applied. Ions are propelled axially by a transient direct current voltage that is applied to each electrode in succession (i.e., the traveling wave).[24, 25] Ions with short drift times traverse the ion guide while being 'carried' by the wave. In contrast, ions with long drift times 'roll over' the wave.[25]

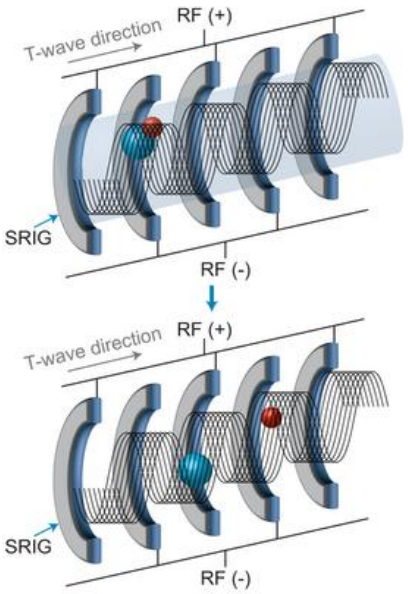

Figure 5.1: TWIMS schematic. A traveling voltage wave (T-wave) applied to a series of electrically connected ring electrodes (stacked ring ion guide: SRIG) pushes ions through the device. For a given wave speed and magnitude, ions carried forward by the T-wave have short drift times (red), while ions that roll over the T-wave take longer to exit the device (blue). RF, radiofrequency. Reprinted with permission from ref 25. Copyright © 2014 Macmillan Publishers Limited.

5.3 Experimental Details

5.3.1 Transmission Mode Direct Analysis in Real Time Metabolic Fingerprinting Method Development

5.3.1.1 Chemicals

D-*erythro*-sphingosine was obtained from Avanti Polar Lipids, Inc. (Alabaster, AL, USA). Anhydrous pyridine and acetone were purchased from EMD Chemicals (Gibbstown, NJ, USA). Polyethylene glycol standard 400 and 600 (PEG 400, 600) were

purchased from Fluka Chemical Corp. (Milwaukee, WI, USA). Healthy human serum (S7023-50 mL) and quinine were procured from Sigma-Aldrich Corp. (St. Louis, MO, USA). Analytical grade isopropanol and *N*-trimethylsilyl-*N*-methylfluoroacetamide (MSTFA) containing 0.1% trimethylchlorosilane (TMCS) were acquired from Thermo Fisher Scientific, Inc. (Waltham, MA, USA). Ultra high purity helium (99.999 %) was purchased from Airgas, Inc. (Atlanta, GA, USA).

5.3.1.2 Serum Sample Preparation

Prior to analysis, healthy human serum samples were taken from a -80 °C freezer and thawed on ice. Two hundred µL serum aliquots were mixed with 1 mL of a chilled (-18 °C) and degassed 2:1 (v/v) acetone/isopropanol mixture, vortex-mixed, and stored overnight at -18 °C to precipitate proteins. Centrifugation of samples at 13,000 x *g* for 5 min followed. The supernatant was transferred to a new centrifuge tube, and the solvent was evaporated in a Savant DNA 110 speed vacuum concentrator. The remaining solid residue was dissolved in 25 µL of anhydrous pyridine in a N_2-purged glove box and shaken for 1 h at room temperature for complete dissolution. Fifty µL of MSTFA containing 0.1% TMCS were added to samples in the glove box. The mixture was incubated at 55 °C for 1 h, resulting in trimethylsilane derivatization of amide, amine, carboxyl, and hydroxyl groups. The derivatized serum samples were subjected to DART-MS analysis as described in Section 5.3.1.4.

5.3.1.3 Mass Spectrometry Instrumentation

MS analysis was performed with a DART SVP 100 source (IonSense, Saugus, MA, USA) coupled to a hybrid QTOF mass spectrometer (Bruker micrOTOF-Q II;

Bremen, Germany). Prior to coupling, the mass spectrometer interface was modified with the addition of a custom-built gas-ion separator tube (GIST) connected to a Vacuubrand 2C diaphragm pump (Vacuubrand, Wertheim, Germany) to reduce the amount of discharge gas flowing into the mass spectrometer atmospheric pressure inlet. A DART helium flow rate of 1 L min^{-1} heated to 250 °C was chosen for all experiments unless otherwise indicated. The exit grid voltage was set at +350 V. The settings for the QTOF mass spectrometer were as follows: inlet capillary voltage, -1500 V, end plate offset, -500 V, drying gas (nitrogen) flow rate, 2.0 L min^{-1}, drying gas temperature, 150 °C, ion funnel 1 RF, 150 V_{pp}, ion funnel 2 RF, 200 V_{pp}, hexapole RF, 100 V_{pp}, quadrupole ion energy, 5 eV, collision cell transfer time, 77 µs, collision cell pre pulse storage, 5 µs. All experiments were conducted in positive-ion mode. Spectra were acquired at 1.0 Hz in the 35 – 1000 m/*z* range. Tandem MS experiments were conducted utilizing nitrogen as the collision gas with 10 and 30 eV collision energies. The instrument was mass calibrated before conducting experiments using a 1:1 solution (v/v) of PEG 400 and 600. Mass spectral data processing, signal averaging, and background subtraction were performed using DataAnalysis 4.0, the built-in mass spectrometer software. SmartFormula and SmartFormula 3D, DataAnalysis 4.0 modules, were used to attain elemental formulas for MS and MS/MS experiments.

5.3.1.4 Direct Analysis in Real Time Setup

For PM-DART analysis, 0.8–1.1 o.d. x 90 mm-long glass capillary probes (Kimble Chase, Vineland, NJ, USA) were used. One µL of sample was deposited on the closed end of the glass probes. Probes were placed on a custom built holder (“module”) mounted on a software-controlled linear rail (IonSense, Saugus, MA, USA) positioned

between the DART ion source and the GIST (Figure 5.2). An automated DART software method was utilized for sample introduction. This method had a waiting period of 5 s, after which the probe was moved towards the ionization region at 10 mm s^{-1}. Before entering into the ionization region, the movement of the probe was paused. Following a speed change to 0.4 mm s^{-1}, the probe was moved an additional 7 mm and was stopped in front of the DART source for data acquisition. The DART software method was started following sample deposition on the probe.

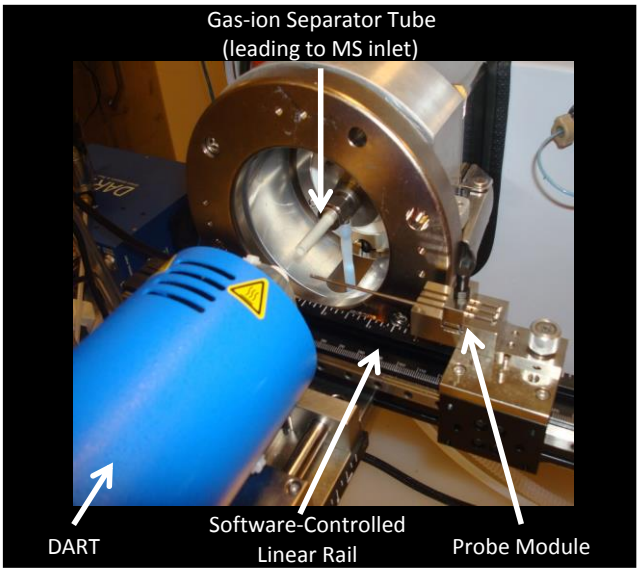

Figure 5.2: Photograph of the automated PM-DART MS system.

Stainless steel (SS) mesh type 304 (IonSense) and type 316 (TWP Inc., Berkeley, CA, USA) were employed as sample substrates for TM-DART experiments. Samples were deposited on designated areas of a clean SS mesh strip (SS T304) or discs (SS

T316). The SS mesh substrates had the following characteristics: wire diameter, 0.004 in (SS mesh strips) or 0.0075 in (SS mesh discs), mesh size, 74 x 74 wires in^{-1} (SS mesh strips) or 60 x 60 wires in^{-1} (SS mesh discs), open area, 52.7% (SS mesh strips) or 30% (SS mesh discs). A 1 in long SS mesh strip or individual 1 cm (diameter) SS mesh discs were placed in the mesh holder module having 10 open slots available for sample deposition (Figure 5.3). The exposed mesh areas in these slots were 33 mm^2. No sample was deposited on the first slot, followed by deposition of 2 µL of derivatized serum solution on the central point of the exposed mesh in the second slot. Deposition of subsequent samples ensued in this alternating manner.

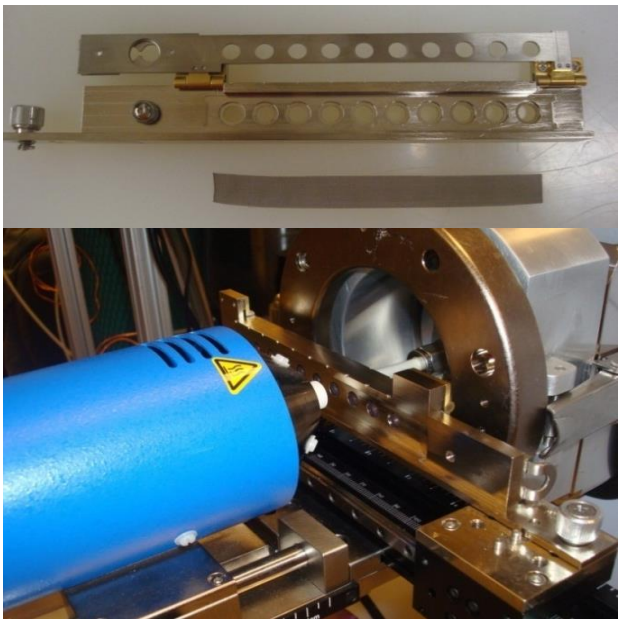

Figure 5.3: Photograph of the automated TM-DART system. The top panel displays the module used to hold the stainless steel mesh strip or discs, as well as a stainless steel mesh strip. The bottom panel shows the automated TM system during an analysis placed between the DART ion source and the GIST inlet.

As with PM-DART, an automated DART software method was used. This method had no waiting period. The module was moved towards the ionization region at 10 mm s^{-1} and stopped when the first exposed mesh was in front of the DART source. The module advanced to adjacent exposed areas at 10 mm s^{-1}.

For all data shown, mass spectrometric signals were only averaged for the first 45 s of the analysis time and experiments were carried out in triplicate. Optical microscope images of samples spotted on meshes were acquired with an Olympus BX41 laboratory microscope equipped with an Olympus DP12 digital microscope camera (Olympus America Inc., Melville, NY, USA). A FLIR T-Series Infrared Camera (FLIR Systems, Inc., Boston, MA, USA) was used to acquire images of the TM-DART experimental setup.

The identity of metabolites was tentatively annotated by matching accurate masses against a custom built database encompassing entries for endogenous human metabolites found in the Human Metabolome Database (HMDB).[26] Entries for compounds reacting with the MSTFA/TMCS reagent mixture were manually expanded to take into account mono-, di-, tri-, tetra-, penta-, and/or hexa-trimethylsilane (TMS) derivatives. Matching metabolites not found in human blood sera were excluded. The matching of experimental data to database accurate mass records was performed using the SearchFromList application included in the Mass Spec Tools suite of programs (ChemSW, Fairfield, CA, USA) using a tolerance of 10 mDa and a relative intensity cutoff of 1%.

5.3.2 Application of Transmission Mode Direct Analysis in Real Time for Cystic Fibrosis Detection

5.3.2.1 Chemicals

Omnisolv LC-MS grade acetonitrile was purchased from EMD (Billerica, MA, USA). Ultrapure water with 18.2 MΩ cm resistivity (Thermo Scientific Barnstead Nanopure UV ultrapure water system, Marietta, OH, USA) was used in sample preparation protocols.

5.3.2.2 Exhaled Breath Condensate Sample Collection and Preparation

EBC samples from 5 healthy controls and 4 patients diagnosed with cystic fibrosis (CF) were collected under the guidelines approved by the Georgia Institute of Technology and the Emory University Institutional Review Boards (approval number IRB00000372). EBC was collected using an R-Tube collector (Respiratory Research, Inc., Austin, TX, USA) and kept frozen at -80 °C until processed. Prior to analysis, EBC samples were thawed, and 2 mL aliquots were placed into separate vials, stored at -80 °C for a minimum of 2 h, and freeze-dried overnight using a VirTis bench top freeze-dryer (SP Industries, Stone Ridge, NY, USA). The remaining residues were reconstituted, without derivatization, in 100 μL of acetonitrile:water 80:20 v/v (concentration factor = 20) and analyzed by TM-DART-TWIMS-TOF MS in the negative ion mode to test the feasibility of this technique in metabolomics investigations.

5.3.2.3 Traveling Wave Ion Mobility Mass Spectrometry Instrumentation

TWIMS-TOF MS was performed in negative ion mode on a Synapt G2 High Definition Mass Spectrometry system (Waters Corporation, Manchester, UK), a hybrid

quadrupole-ion mobility-orthogonal acceleration TOF mass spectrometer with a typical resolving power of 20,000 m/Δ*m* (FWHM) and mass accuracy of 9 ppm at m/*z* 554.2615. TWIMS parameters were as follows: wave height ramped between 10–40 V; wave velocity ramped between 400–1000 m s^{-1}, IMS gas flow rate: 40 mL min^{-1}; helium gas flow rate: 180 mL min^{-1}. The mass spectrometer was calibrated across the range of m/*z* 50–1000 using a 0.5 mM sodium formate solution prepared in 90:10 2-propanol:water v/v. Raw data were evaluated using MassLynx v4.1.

5.3.2.4 Transmission Mode Direct Analysis in Real Time Setup

A DART SVP 100 ion source and transmission module (IonSense, Saugus, MA) were used to conduct TM-DART experiments at a 1 L min^{-1} He gas flow rate. The discharge gas was heated to 300 °C. A custom-built flange and gas-ion separator tube (GIST), connected to a Vacuubrand 2C diaphragm pump (Vacuubrand, Wertheim, Germany), was used to couple the DART ion source to the mass spectrometer interface to reduce the amount of discharge gas flowing into the mass spectrometer atmospheric pressure inlet. The exit grid voltage was set at 300 V. Individual 1 cm stainless steel mesh discs were placed within the 10-position transmission module and 4 µL of sample (i.e. solvent blank or EBC concentrate) was deposited in the center of the exposed mesh area and allowed to dry for approximately 5 min. Three technical replicates were analyzed for each sample, leaving a blank mesh positioned before every deposited sample spot. An automated method was used to introduce the transmission module into the DART ionization region at 10 mm s^{-1}, and each position was held within the DART ionization region for 2 min. To avoid disturbing the DART ionizing gas stream, only a 1 s hold time was used when the transmission module was advanced to the next position.

5.3.2.5 Data Analysis

Metabolic features were extracted from TWIMS-TOF MS data as drift time-m/z pairs using MZmine 2.10.[27] Although originally designed for extracting chromatographic data, the drift time function of the acquired TWIMS-TOF MS data could be analyzed similarly using MZmine 2.10, as the ion mobility chronograms obtained from DART experiments resemble typical chromatograms. The feature extraction workflow consisted of mass detection followed by the generation, deconvolution, de-isotoping, alignment, and integration of extracted ion mobility chronogram peaks. Due to the nature of ion separations in IM-MS experiments, where different adduct species do not necessarily coincide in drift time, mass spectral adduct ion analysis was not performed. EBC sample data were curated against corresponding solvent blank data to determine EBC-specific features, where features with peak areas ≥10× those present in solvent blanks were deemed EBC-specific.

The resulting matrix from metabolic spectral feature extraction of TM-DART TWIMS-TOF MS data from 5 healthy controls and 4 CF patients was utilized to build a model for sample class discrimination via oPLS-DA and to down-select a smaller panel of highly-discriminant features through the usage of a genetic algorithm (MATLAB Version 7.13.0, The MathWorks, Inc., Natick, MA, USA with PLS_Toolbox v.6.71, Eigenvector Research, Inc., Wenatchee, WA, USA). A panel of 9 discriminant features had the lowest root-mean-square error of cross-validation (RMSECV) at the conclusion of the genetic algorithm variable selection process. The parameters for genetic algorithm variable selection were as follows: population size: 64, variable window width: 1, % initial terms (variables): 15, target minimum # of variables: 1, target maximum # of

variables: 8, penalty slope: 0.005, maximum generations: 150, % at convergence: 79.7, mutation rate: 0.005, crossover: double, regression choice: PLS, # of latent variables: 4, cross-validation: random, # of splits: 3, # of iterations: 5, replicate runs: 10. PLS-DA models were orthogonalized and internally cross-validated using leave-one-out cross-validation. Data were preprocessed by mean-centering prior to oPLS-DA analysis. Principal component analysis (PCA) was also performed to inspect data before and after genetic algorithm variable selection (i.e., on all of the extracted spectral features and only the discriminant feature panel).

MetaboSearch[28] was utilized to search the METLIN database,[29] the LIPID Metabolites and Pathways Strategy (LIPID MAPS) database,[30] and the human metabolome database (HMDB)[26] for endogenous metabolite candidates solely using neutral masses with a mass accuracy error no greater than 15 mDa.

5.4 Optimization of Transmission Mode Direct Analysis in Real Time Experimental Variables

A typical TM-DART mass spectrum of MSTFA-derivatized human blood serum, along with microscopic images of the sample spot before and after analysis, is shown in Figure 5.4.

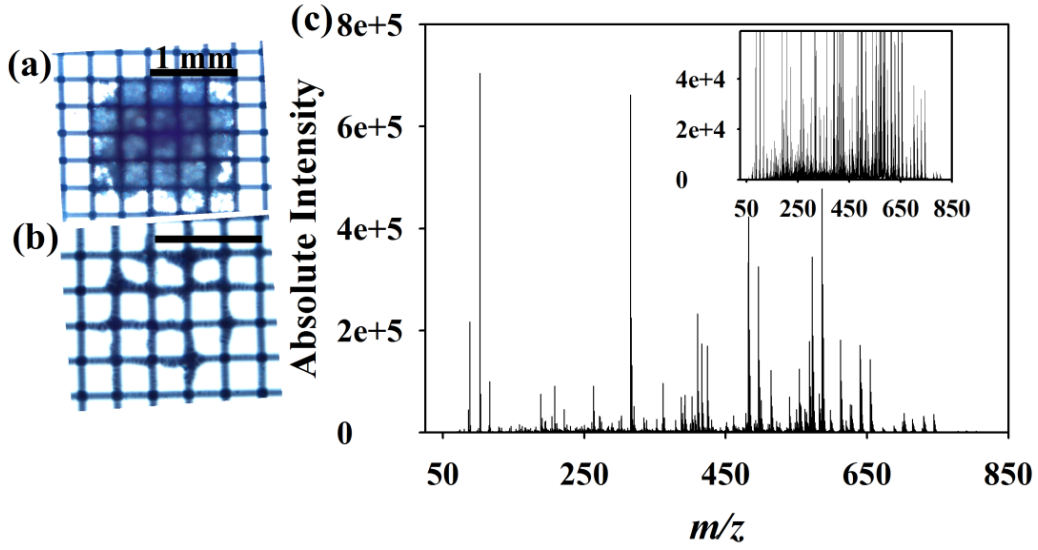

Figure 5.4: Microscopic image of a dried, derivatized human serum sample (a) before and (b) after TM-DART analysis with the (c) corresponding background-corrected positive ion mode mass spectrum in the 50 – 850 m/*z* range. The inset details the signals observed upon zooming into the baseline.

It was observed that, once dried, samples "sat" on the SS mesh instead of being interspersed around the SS strands which constitute the mesh (Figure 5.4a). Exposure to the heated DART stream resulted in desorption of the entire sample (Figure 5.4b) after 45 seconds. The ensuing mass spectrum was highly complex (Figure 5.4c) as many metabolite-related features were detected. Closer investigation of this TM-DART mass spectrum (Figure 5.4c, inset) revealed that many less-dominant ion signals also existed. Table 5.1 shows elemental formulas assigned to species present in the spectrum shown in Figure 5.4, and their tentative annotations based on matches to entries in the HMDB.

An investigation of the sources of sensitivity drift in TM-DART was carried out to better understand what experimental variables should be controlled for ensuring

optimum coefficients of variation (CV) values. In initial exploratory experiments using SS mesh strips, a progressive increase in the total ion current (TIC) was observed for repeat analyses (Figure 5.5).

Table 5.1. Tentative annotation of peaks selected from the chemically derivatized human blood serum mass spectrum shown in Figure 5.2 via accurate mass measurements with a mass tolerance of 10 mDa and a relative intensity cutoff of 1% of the base peak.

Measured Ions *(m/z)*	Ion Type	Experimental Exact Mass (Da)	Theoretical Exact Mass (Da)	Accuracy (mDa)	Estimated Formula	Name	Source
146.1004	$[M+H]^+$	145.0926	145.0851	7.5	$C_5H_{11}N_3O_2$	4-Guanidinobutanoic acid	HMDB03464
190.1289	$[M+H]^+$ $[M+2TMS+H]^+$	189.1211	189.1113	9.8	$C_7H_{15}N_3O_3$	Homocitrulline	HMDB00679
204.1434	$[M+2TMS+H]^+$	59.0565	59.0483	8.2	CH_5N_3	Guanidine	HMDB01842
221.1288	$[M+TMS+H]^+$	148.0814	148.0735	7.9	$C_6H_{12}O_4$	Mevalonic acid	HMDB00227[a]
250.1344	$[M+TMS+H]^+$	177.0870	177.0789	8.1	$C_{10}H_{11}NO_2$	5-Hydroxytryptophol	HMDB01855[b]
263.1405	$[M+2TMS+H]^+$	118.0536	118.0629	9.3	$C_5H_{10}O_3$	3-Hydroxyvaleric acid	HMDB00531[c]
289.1380	$[M+2TMS+H]^+$	144.0511	144.0422	8.9	$C_6H_8O_4$	3-Methylglutaconic acid	HMDB02266[d]
320.1738	$[M+2TMS+H]^+$ $[M+3TMS+H]^+$ $[M+5TMS+H]^+$	175.0869	175.0956	8.7	$C_6H_{13}N_3O_3$	Citrulline	HMDB00904[e]
334.1900	$[M+3TMS+H]^+$ $[M+4TMS+H]^+$ $[M+5TMS+H]^+$	117.0636	117.0538	9.8	$C_3H_7N_3O_2$	Guanidoacetic acid	HMDB00128

Table 5.1 (continued).

442.2507	$[M+4TMS+H]^+$	153.0848	153.0789	5.9	$C_8H_{11}NO_2$	Dopamine	HMDB00073[f]
461.1880	$[M+3TMS+H]^+$	244.0616	244.0695	7.9	$C_9H_{12}N_2O_6$	Uridine	HMDB00296[g]
461.1880	$[M+4TMS+H]^+$	172.0220	172.0136	8.4	$C_3H_9O_6P$	Glycerol 3-phosphate	HMDB00126[h]
468.2395	$[M+4TMS+H]^+$	179.0735	179.0793	5.8	$C_6H_{13}NO_5$	Glucosamine	HMDB01514[i]
492.2657	$[M+3TMS+H]^+$	275.1392	275.1481	8.8	$C_{11}H_{21}N_3O_5$	Epsilon-(gamma-Glutamyl)-lysine	HMDB03869[j]
512.2772	$[M+TMS+H]^+$ $[M+2TMS+H]^+$ $[M+4TMS+H]^+$	439.2299	439.2392	9.3	$C_{23}H_{37}NO_5S$	Leukotriene E4	HMDB02200
514.2897	$[M+H]^+$ $[M+TMS+H]^+$ $[M+2TMS+H]^+$	513.2819	513.2760	5.9	$C_{26}H_{43}NO_7S$	Sulfolithocholylglycine	HMDB02639
526.2593	$[M+4TMS+H]^+$	237.0934	237.0861	7.2	$C_9H_{11}N_5O_3$	Biopterin	HMDB00468[k]
700.3462	$[M+2TMS+H]^+$	555.2593	555.2692	9.9	$C_{28}H_{37}N_5O_7$	Enkephalin L	HMDB01045
700.3462	$[M+6TMS+H]^+$	267.1012	267.0954	5.8	$C_9H_{17}NO_8$	Neuraminic acid	HMDB00830

Table 5.1 (continued).

[a]1 isomer found with source ID HMDB12140; [b]1 isomer found with source ID HMDB12490; [c]10 isomers found with source IDs HMDB02011, HMDB00351, HMDB00410, HMDB00642, HMDB01863, HMDB00354, HMDB00396, HMDB00407, HMDB01987, HMDB00754; [d]3 isomers found with source IDs HMDB13311, HMDB02266, HMDB00393; [e]1 isomer found with source ID HMDB03148; [f]1 isomer found with source ID HMDB04825; [g]1 isomer found with source ID HMDB00767; [h]1 isomer with source ID HMDB02520; [i]1 isomer found with source ID HMDB02030; [j]1isomer found with source ID HMDB04207; [k]5 isomers found with source IDs HMDB01195, HMDB02263, HMDB00633, HMDB00817, HMDB00238.

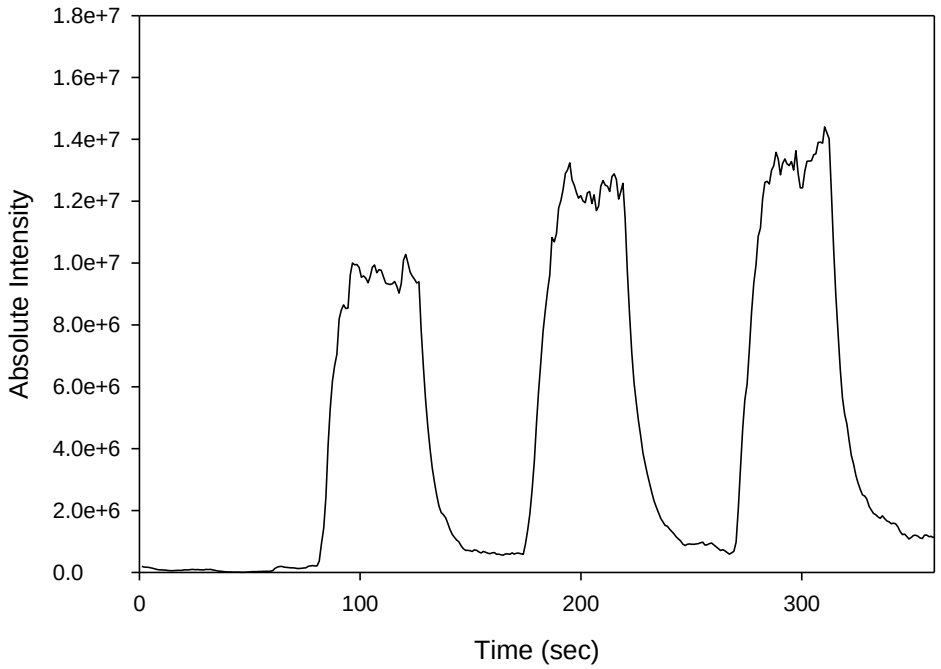

Figure 5.5: Total ion chronogram observed during TM-DART analyses of a derivatized serum sample showing an increase in the abundance of detected ions for each successive analysis prior to sample introduction method modification.

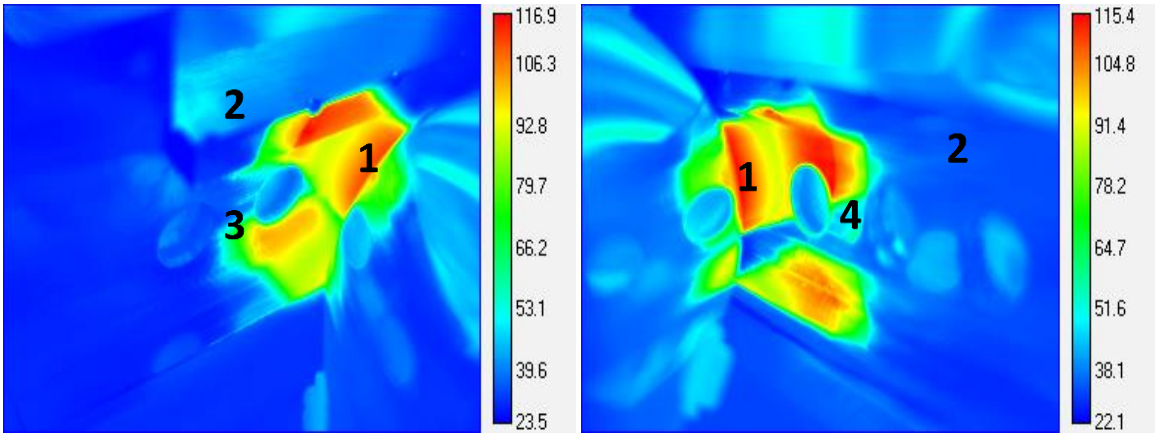

Figure 5.6: Thermal IR images of the TM-DART ion source and sampling module assembly during the analysis of a derivatized serum samples. The DART cap (1) in the IR images rests in close proximity of the module (2) where the sample is spotted for analysis. Two previously-analyzed mesh holder samples positions (3) are shown on the left, and two mesh positions in the queue (4), waiting to be analyzed, on the right. The color scale displays the measure temperature in Celsius. The set DART temperature was 250 °C.

As metabolite ion intensities were monitored, this phenomenon greatly impacted the reproducibility of TM-DART analyses. To investigate possible origins of this temporal drift effect, the TM-DART experimental set-up was thermally imaged with an IR camera during the analysis of a serum sample (Figure 5.6). A temperature increase in parts of the TM module surrounding the spot being analyzed was observed. Additionally, the sections of the SS mesh strip exposed in the two spots adjacent to the spot being sampled were also observed to be heated to higher-than-ambient temperature (~52°C) due to the good thermal conductivity of stainless steel. It is likely that this heat gain could allow for some extent of sample carryover (or loss) from one sampled spot to the next one during TM-DART analyses. Moreover, fringe heating of a sample prior to its analysis could cause premature melting, leading to a subsequent redistribution of the sample on the mesh. It is also plausible that desorption and ionization processes would be more efficient on a "pre-heated" sample, causing a progressive increase in signal as samples placed later in the analysis queue were interrogated. In an effort to mitigate these potential thermal effects, SS mesh discs instead of strips were used for TM experiments. As an additional cautionary measure, samples were deposited in an alternating manner, leaving one SS mesh disc void of sample between each SS mesh disc being analyzed.

As the key step in the DART ionization mechanism involves proton transfer reactions from protonated water cluster reagent ions to desorbed analyte neutrals, the intensities of various reagent ions $(H_2O)_nH^+$ (n = 5, 6, 8, 11, 13) were monitored to determine if these species were consistently available during successive same sample analyses (Figure 5.7). Data was collected immediately after the DART corona-glow discharge was activated. As shown, the intensities of the monitored reagent ions

increased for at least 30 minutes before stabilizing, despite the fact that the temperature of the ionization region stabilized within 1 minute and the relative humidity in the environment was constant. Therefore, part of the observed increase in TIC for repeat serum sample analyses during this "warm up" period seemed to correlate with observed changes in reactant ion abundances, indicating that, under these conditions, a finite time is needed for species in the DART plasma and/or ionization region to reach a steady state which would ensure that subsequent experiments are reproducible. Because proton transfer reactions of basic polar molecules under atmospheric pressure conditions are under kinetic control, following pseudo-first order reaction kinetics, changes in the reactant ion population are bound to change reaction rate constants. Upon investigation, it was found that the original vendor TM-DART software methods deactivated and reactivated the DART corona-glow discharge prior to the TM module entering into the ionization region. Consequently, any efforts to maintain reagent ion signal stability using a "warm up" period were negated. New software methods that did not deactivate the DART corona-glow discharge were subsequently created to solve this problem. The aforementioned implementation of SS mesh discs in TM-DART experiments, modification of the software sample insertion methods for ensuring stabilization of the reagent ion population, and leaving SS mesh discs void of sample between serum samples, improved CV values for monitored detected metabolite ions by as much as 50%.

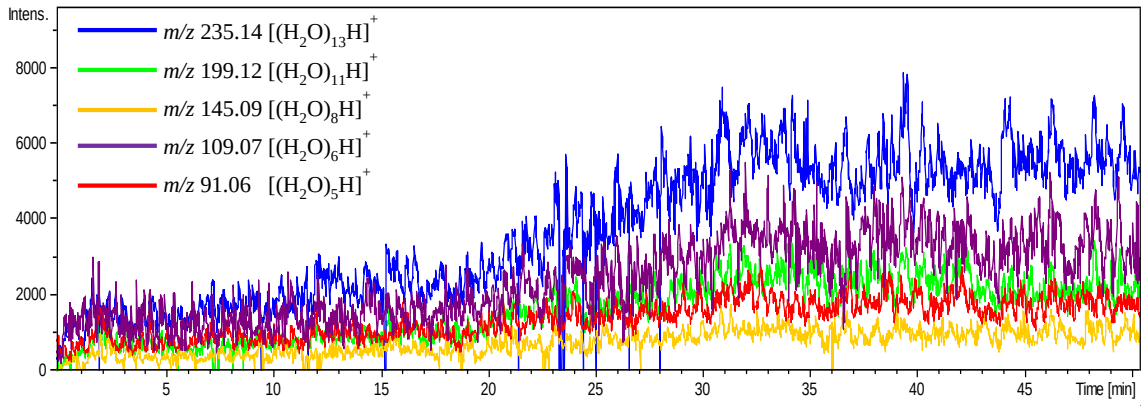

Figure 5.7: Extracted ion chronograms for selected reagent water cluster ions present in the DART ionization regions for a period of ~50 minutes after application of high voltage to the discharge electrode. No signals for reagent ions with m/*z* lower than ~75 were observed.

As a primary aim of untargeted metabolic profiling is to acquire maximum coverage for detected metabolites both in terms of sensitivity and mass range, the key DART experimental variables affecting metabolite desorption and ionization were optimized. First, the sensitivity and reproducibility of TM-DART MS for liquid and dried blood serum samples were investigated via 3 sample deposition/introduction strategies: (A) serum sample introduced into the ionization region without prior drying, (B) serum sample dried with heated nitrogen from the DART source with the discharge deactivated, and (C) serum sample self-dried on mesh before introduction into the ionization region (Figure 5.8). For the monitored ionic species (m/*z* 315.1042, $C_{13}H_{23}O_5Si_2^{\cdot}$, m/*z* 416.2246, $C_{19}H_{38}NO_5Si_2^{\cdot}$, m/*z* 640.3391, $C_{33}H_{54}N_3O_4Si_3^{\cdot}$, m/*z* 714.3601, $C_{29}H_{60}N_7O_6Si_4^{\cdot}$), the sample deposition methods that involved a drying step before analysis (B and C) had CVs of 30% and 18.3%, respectively. Utilizing a non-dried serum sample resulted in less precise results (CV: 26.3%) than the self-dried serum sample.

Interestingly, the intensities of the lower mass ionic species ($m/z < 600$) were higher when samples were first dried, while the higher mass species ($m/z > 600$) showed an inverse behavior. Although this observation is not yet completely understood, a "splattering" effect may be the probable cause of this observed trend. The non-dried serum sample may begin to splatter when exposed to the DART gas stream, so a loss of the more volatile, lower mass species could more easily occur as a result of thermal losses during the desorption process. Higher mass species may benefit from desorption from a "wet" sample, as the solvent removes intermolecular interactions which reduce volatility in the solid-phase.

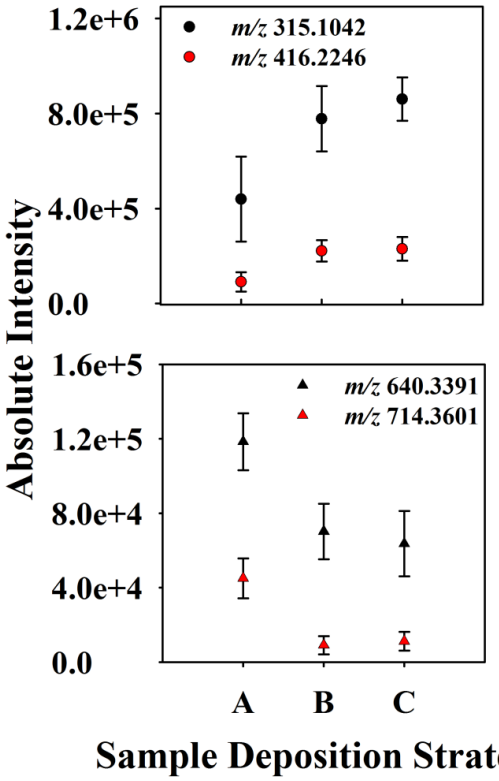

Figure 5.8: Effect of serum sample deposition strategy (A: liquid sample; B: sample dried with heated N_2 from DART source; C: self-dried sample) on sensitivity and reproducibility of successive TM-DART analyses for mass spectrometric signals at m/z = 315.1042, $C_{13}H_{23}O_5Si_2$, m/z = 416.2246, $C_{19}H_{38}NO_5Si_2$, (top panel) and signals at m/z = 640.3391, $C_{33}H_{54}N_3O_4Si_3$, m/z = 714.3601, $C_{29}H_{60}N_7O_6Si_4$, (bottom panel).

Experiments also showed that the position of the DART gas outlet relative to the SS mesh greatly affects the intensity and information content of the observed signal from serum samples analyzed by TM-DART MS (Figure 5.9). Sensitivity was greatly reduced when the DART gas outlet was not in close proximity of the SS mesh holder (ii > 1 mm in Figure 5.9a). This effect is most likely caused by excessive cooling leading to poor thermal desorption. Therefore, only the effect of mesh-to-GIST distance (indicated as “i”

in Figure 5.9a) on sensitivity was investigated. The current TM-DART set-up limits mesh-GIST distances to 5, 12, and 19 mm.

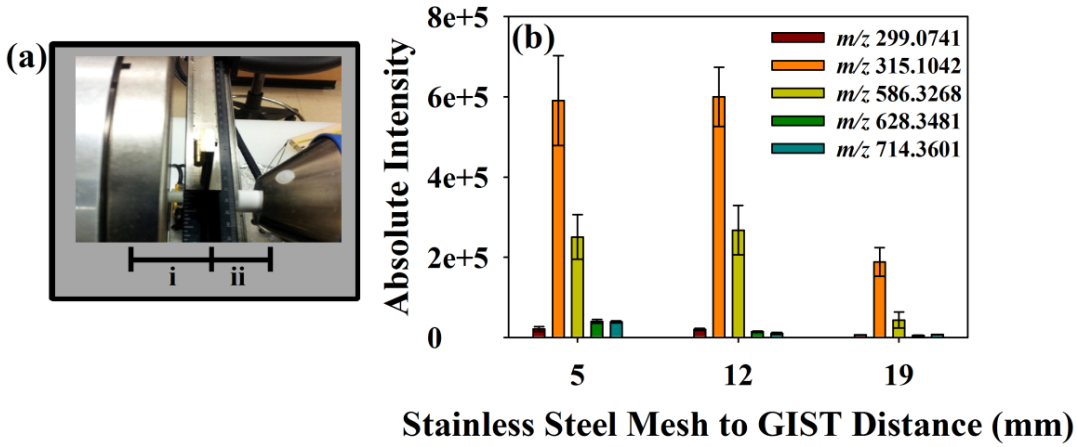

Figure 5.9: The sensitivity of TM-DART MS analysis is influenced by the position of the SS mesh substrate in the ionization region. Two parameters (a) determine this position: the distance from the GIST to the SS mesh (i) and the distance from the mesh to the DART cap (ii). The effect of varying distance "i" on the sensitivity for untargeted metabolic profiling of derivatized serum is displayed in (b).

Higher intensities for metabolite ions were obtained at mesh-GIST distances of 5 and 12 mm. Coverage for detected metabolites extended to 800 m/z for these distances. A major decrease in intensity occurred at 19 mm and not many metabolite ions beyond 700 m/z were detected. The shorter mesh-GIST distances probably allowed for enhanced fluid dynamic focusing of the ions toward the GIST, as previously shown by fluid dynamic simulations.[10] These simulations showed that in TM-DART, the heated gas plume is

channeled almost completely through the sample area when the DART cap is in close proximity to the module. Since this is also likely occurring in the current system, the farther away the mesh is positioned from the GIST, the poorer the ion transmission due to divergence of the gas stream and the resultant observed sensitivity.

Figure 5.10a demonstrates the effect of increasing the set DART temperature on the intensity of lower mass species (m/z < 600). Overall, a slightly curved-shaped trend was observed.

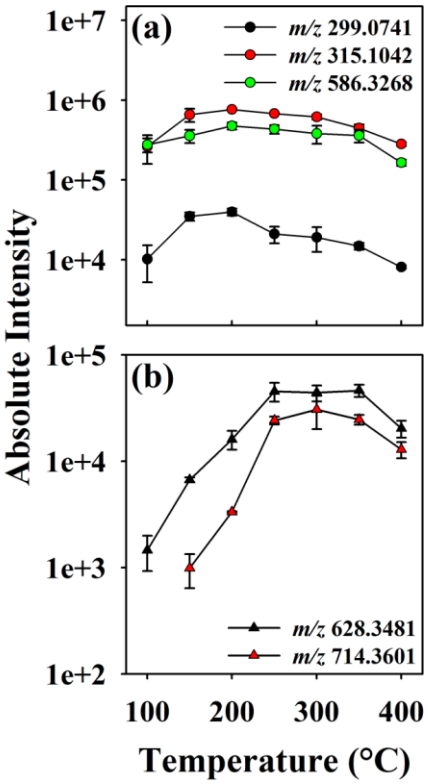

Figure 5.10: Effect of helium gas temperature on TM-DART MS sensitivity for untargeted metabolic profiling of derivatized serum: (a) absolute intensities of monitored signals in the low mass range and (b) absolute intensities of monitored signals at higher masses. Intensities were monitored at DART set temperatures of 100, 150, 200, 250, 300, 350, and 400 °C.

Lower temperatures were efficient enough to desorb lower mass metabolites from samples with the best sensitivity. Increasing the temperature beyond 200 °C conceivably allowed for thermal degradation of low mass metabolites.[31] Conversely, the intensity of higher-mass ionic species ($m/z > 600$) increased with increasing temperature before beginning to stabilize at 300 °C (Figure 5.10b). Higher temperatures resulted in more efficient desorption of higher mass metabolites as the amount of thermal energy required for their volatilization is higher compared to that of lower mass metabolites. Increasing the temperature beyond 350 °C most likely increased the thermal degradation rate of all metabolites.

An advantage of hyphenated platforms for metabolic profiling is the reduced potential for ion suppression effects. As the DART analyte desorption process is thermal, temperature ramping can be used to partially separate metabolites in time based on their volatility.[32, 33] A pseudo-separation dimension can be incorporated to the analyses while maintaining the high throughput nature of DART. Figure 5.11 shows extracted ion chronograms for several ionic species produced from a blood serum sample for a three minute temperature ramping experiment starting at 150 °C and ending at 450 °C. Varying desorption profiles were observed for the ionic species. Generally, the intensities of lower mass metabolic ionic species began to decay after 30 seconds, while intensities of higher mass metabolic ionic species continuously increased for two minutes.

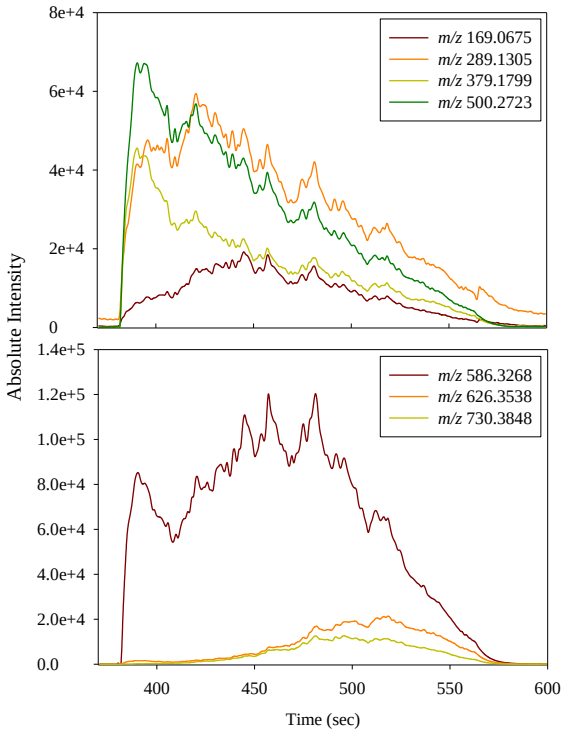

Figure 5.11: Effect of ramping the set helium gas temperature on intensities of monitored mass spectrometric signals in the low (top panel) and high (bottom panel) mass ranges. The set helium gas temperature was ramped from 150 °C to 450 °C over 3 minutes.

Increasing the flow rate of the ionizing gas from 0.5 to 1.0 L min^{-1} resulted in an upward trend for the intensities of all of the monitored metabolic ionic species (Figure 5.12). Increased flow rates resulted in more reactive species interacting with serum samples in a set time period, thereby, allowing for enhanced ionization. As the ionizing gas stream additionally aids in ion transport to the GIST, sensitivity was also enhanced through this secondary effect. Flow rates higher than 1.0 L min^{-1} resulted in slightly decreased intensities for all of the monitored metabolic ionic species and undesirable mass analyzer pressures, so they were not routinely used.

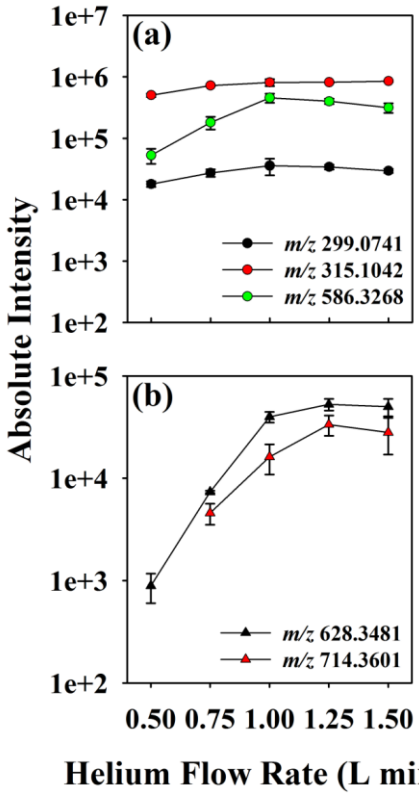

Figure 5.12: Effect of helium gas flow rate on TM-DART MS sensitivity for untargeted metabolic profiling of derivatized serum: (a) absolute intensities for selected ions below m/z = 600, (b) absolute intensities for selected ions above m/z = 600. Intensities were monitored at flow rates of 0.50, 0.75, 1.00, 1.25, and 1.50 L min^{-1}.

5.5 Comparison of Probe Mode Direct Analysis in Real Time and Transmission Mode Direct Analysis in Real Time

The antimalarial quinine was used as a test compound to investigate the precision of both PM- and TM-DART operational modes in the absence of matrix effects. An internal standard was not used due to the difficulties associated with assessing the type of compound that should be utilized when analyzing a large number of unknown metabolites and evidence of the occurrence of major ion suppression when an internal

standard was added to serum samples (data not shown). It was found that CV values for the $[M+H]^+$ quinine ion were lower for TM-DART than PM-DART (Table 5.2), thus supporting the hypothesis that minimal variance in sample positioning, and minimization of turbulence with a flow-through configuration increases repeatability.

Table 5.2: Reproducibility comparison between automated PM- and TM-DART experiments. The absolute intensities of $[M+H]^+$ quinine ions were monitored (n = 3) at concentrations of 10 and 15 µM. No outliers were removed from the datasets.

Absolute Intensity	10 µM		15 µM	
	Average	CV%	Average	CV%
Transmission Mode	$6.2x10^3$	21	$1.1x10^4$	16
Probe Mode	$1.4x10^5$	92	$1.6x10^5$	35

Repeat PM-DART analysis of the 10 µM quinine solution resulted in a CV of 92%. Upon further inspection of the data, an outlier data point was detected. If this outlier was removed from the data set, the CV was reduced to 50%, which was still greater than TM-DART. During this work, we commonly encountered outliers with PM-DART. The smaller diameter of the probes utilized for PM-DART caused the sample deposition process to be more difficult and less reproducible. Manually depositing samples on the same probe position for all experiments is exceedingly difficult, and sample dispersion (wetting) on the probe is often observed. Overall, we found that TM-DART required less technical replicates and post-analysis user intervention than PM-DART.

Quinine calibration curves generated by TM-DART were found to have a better correlation coefficient and dynamic range than PM-DART, but lower sensitivity (Figure 5.13).

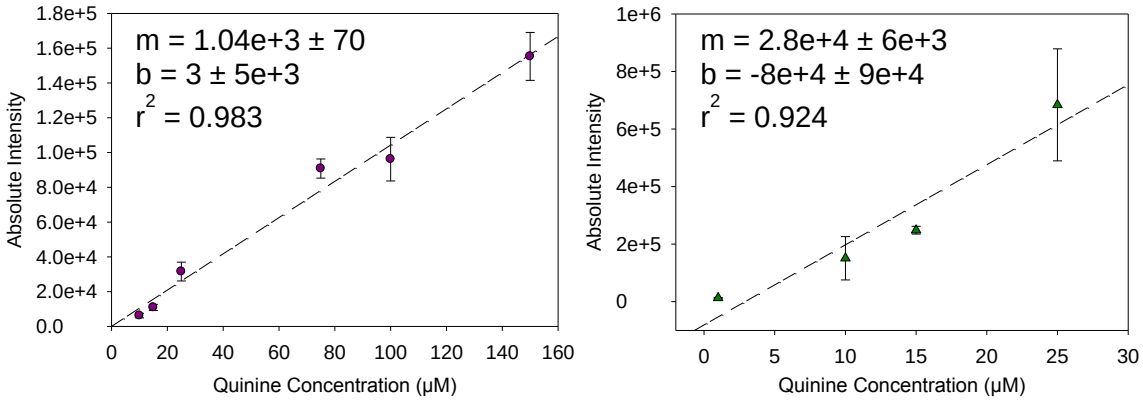

Figure 5.13: Signal linearity for TM-DART (left) and PM-DART (right). The absolute intensity of the $[M+H]^+$ quinine ion was monitored (n = 3) for 1 (PM), 10, 15, 25, 75 (TM), 100 (TM), and 150 (TM) µM solutions. The experimental data was linearly fitted to a $y = mx + b$ model. The regression parameters for each DART operational mode are displayed within each panel.

Due to the thermal desorption processes involved in DART ionization, the sensitivity observed in both modes was directly related to the thermal conductivities of the substrate materials over which the sample was deposited (glass: 1 W m^{-1} K^{-1}, stainless steel: 15 W m^{-1} K^{-1}).[34, 35] During TM-DART analyses, stainless steel more efficiently dissipates the heat from the helium plasma plume, resulting in less thermal energy focused on the sample for analyte desorption. As the glass probes are thermally insulating, the heating

rate for analyte desorption in PM-DART is higher, thus leading to a more transient and efficient desorption profile (Figure 5.14).

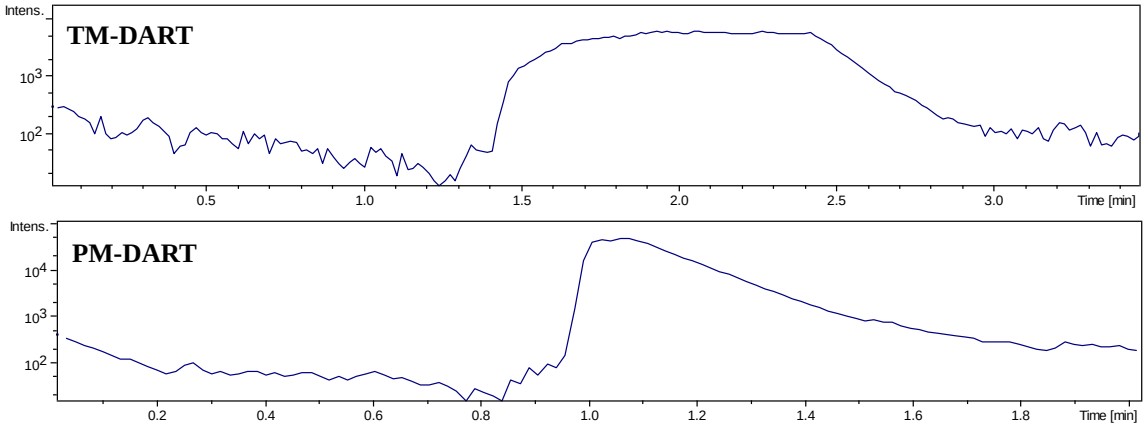

Figure 5.14: Extracted ion chronograms of the $[M+H]^+$ (m/*z* 325.1920) ion observed during TM-DART (top panel) and PM-DART (bottom panel) analysis of a 15 µM quinine solution.

5.6 Transmission Mode Direct Analysis in Real Time Tandem Mass Spectrometry

A distinct aspect of TM-DART is the long-lasting ion signal observed (Figure 5.14), with a duration much greater than that for PM-DART. These long lasting signals allow tandem MS experiments with more constant precursor ion intensity, thus enabling more confident metabolite identification.[36, 37] High mass accuracy product ion spectra not only enable down-selecting likely metabolites from a list of candidates, but also increase the confidence of a given match by determining whether a significant fraction of the fragment ions' elemental formulae are "contained" in the candidate elemental formula of the precursor ion.

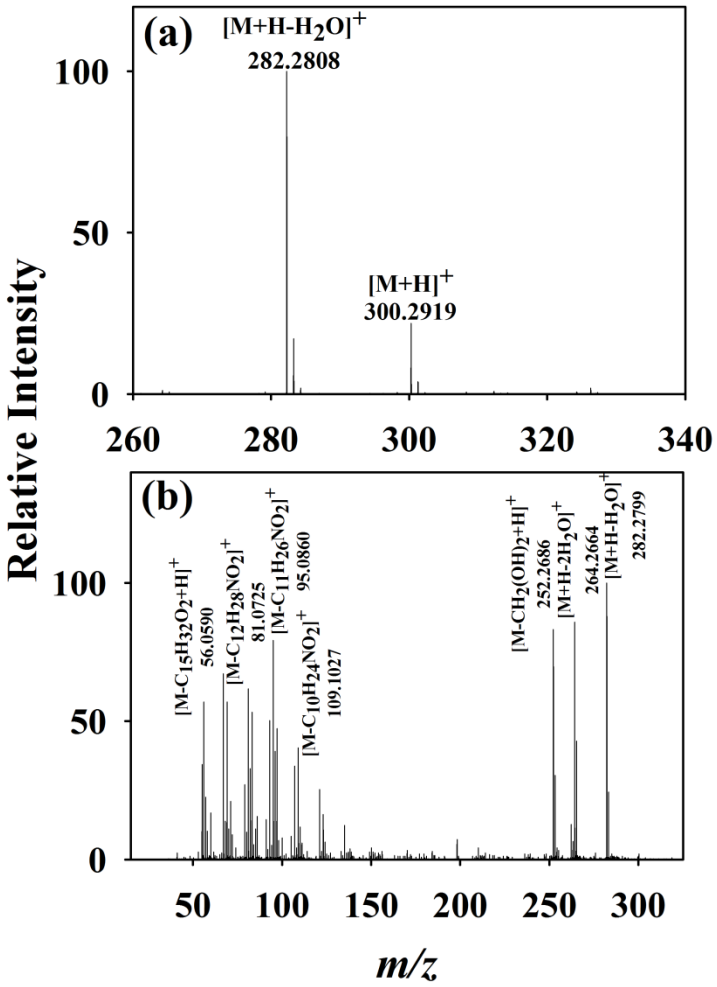

Figure 5.15: TM-DART full scan MS (a) and 30 eV product ion spectrum (b) for D-*erythro*-sphingosine. The quadrupole-selected precursor ion was m/*z* 300.2919.

As an initial assessment of the tandem MS capabilities of the TM-DART approach, we tested the system on a model lipid species. Particularly, sphingolipids are involved in the regulation of cellular proliferation and death through their role as signaling molecules.[38, 39] Due to their structural diversity, many sphingolipids share the same elemental formula. D-*erythro*-sphingosine ($C_{18}H_{37}NO_2$), was fragmented using the TM-DART MS/MS method (Figure 5.15). The $[M+H]^+$ precursor ion derived from the survey spectrum had two elemental formula matches when a tolerance of 15 ppm was used: $C_{18}H_{38}NO_2$ and $C_{16}H_{36}N_4O$ (Table 5.3), with $C_{18}H_{38}NO_2$ having the best score because of the lower mass error.

Table 5.3: Elemental formula result generated by SmartFormula from the full DART MS scan of the $[M+H]^+$ D-*erythro*-sphingosine ion.

Measured m/z: 300.2919				
Formula	Score (%)	m/z	Mass Error (ppm)	mSigma
$C_{18}H_{38}NO_2$	100	300.2897	-7.3	17.1
$C_{16}H_{36}N_4O$	32	300.2883	-11.8	11.1

Incorporating fragmentation data resulted in three elemental formula possibilities within a 15 ppm error (Table 5.4). Both $C_{18}H_{38}NO_2$ and $C_{16}H_{36}N_4O$ explained the same

percentage of the fragment peaks, but $C_{18}H_{38}NO_2$ remained the best match as it had the best mass accuracy (error of -2.6 ppm) and isotopic pattern matching (lowest mSigma value). Incorporating the fragmentation data lowered the mSigma value of $C_{18}H_{38}NO_2$ relative to that of $C_{16}H_{36}N_4O$. In this case, the additional metabolite fragmentation data enabled better down-selection of the correct precursor ion elemental formula candidate. Data of this nature is immensely valuable for metabolomic studies in which the identification of key metabolites is sought after.

Table 5.4: Elemental formula results generated by SmartFormula 3D from the 10 eV MS/MS scan of D-*erythro*-sphingosine. The $[M+H]^+$ ion was identified.

Measured m/z: 300.2904				
Precursor Ion Formula	m/z	Mass Error (ppm)	mSigma	Fragment Peaks Explained (%)
$C_{18}H_{38}NO_2$	300.2897	-2.57	74.2	25.0
$C_{16}H_{36}N_4O$	300.2883	-7.04	79.8	25.0
$C_{14}H_{34}N_7$	300.2870	-11.52	85.8	12.5

5.7 Application of Transmission Mode Direct Analysis in Real Time Metabolic Fingerprinting

TM-DART was interfaced with TWIMS-TOF MS to investigate the feasibility of utilizing TM-DART for metabolomics investigations. Experiments have shown that TM-DART-TWIMS-MS adds to the chemical diversity of metabolites detected in exhaled breath condensate (EBC) (data not shown: experimental data is currently being analyzed by Fernández research group member José J. Pérez), so TM-DART-TWIMS-MS EBC metabolomics was employed for cystic fibrosis (CF) detection. CF is an autosomal recessive disorder caused by a mutated cystic fibrosis transmembrane conductance regulator (CFTR) gene. CFTR is a voltage-gated chloride channel.[40] Metabolomics has the potential to expose metabolic alterations associated with CF pathology and aid in the discovery of biomarkers for which new CF therapeutics can target.[41] TM-DART-TWIMS-TOF MS metabolic profiles were acquired from the exhaled breath condensate (EBC) of 4 CF patients and 5 healthy controls. A total of 66 metabolic features (unique d_t, m/z pairs) were extracted from the acquired metabolic profiles. An oPLS-DA model built from these extracted features was used to discriminate EBC samples from CF patients and healthy controls. Two latent variables interpreting 46.43% of the data matrix variance from the X- (feature peak areas) block and 93.79% of the variance from the Y- (class membership) block resulted in 1 sample misclassification (Figure 5.16a and 5.16b). The cross-validated accuracy, sensitivity, and specificity were 88%, 100%, and 75%, respectively. Although the oPLS-DA model performance was not entirely poor, genetic algorithms were used to attain a smaller, yet more robust, metabolic feature set that could serve to better discriminate between CF patients and normal controls. A set of 9

metabolic features was selected through the genetic algorithm variable selection process as having the lowest RMSECV. The oPLS-DA model generated with this smaller panel (Figure 5.16c and 5.16d) resulted in 100 % cross-validated accuracy, sensitivity, and specificity; hence, no samples were misclassified. This model interpreted 84.24% and 94.29% of the X- and Y-block variances, respectively. The captured X-block variance was 1.8 times greater, thereby, demonstrating that the down-selected set of 9 features was more informative than the initial set.

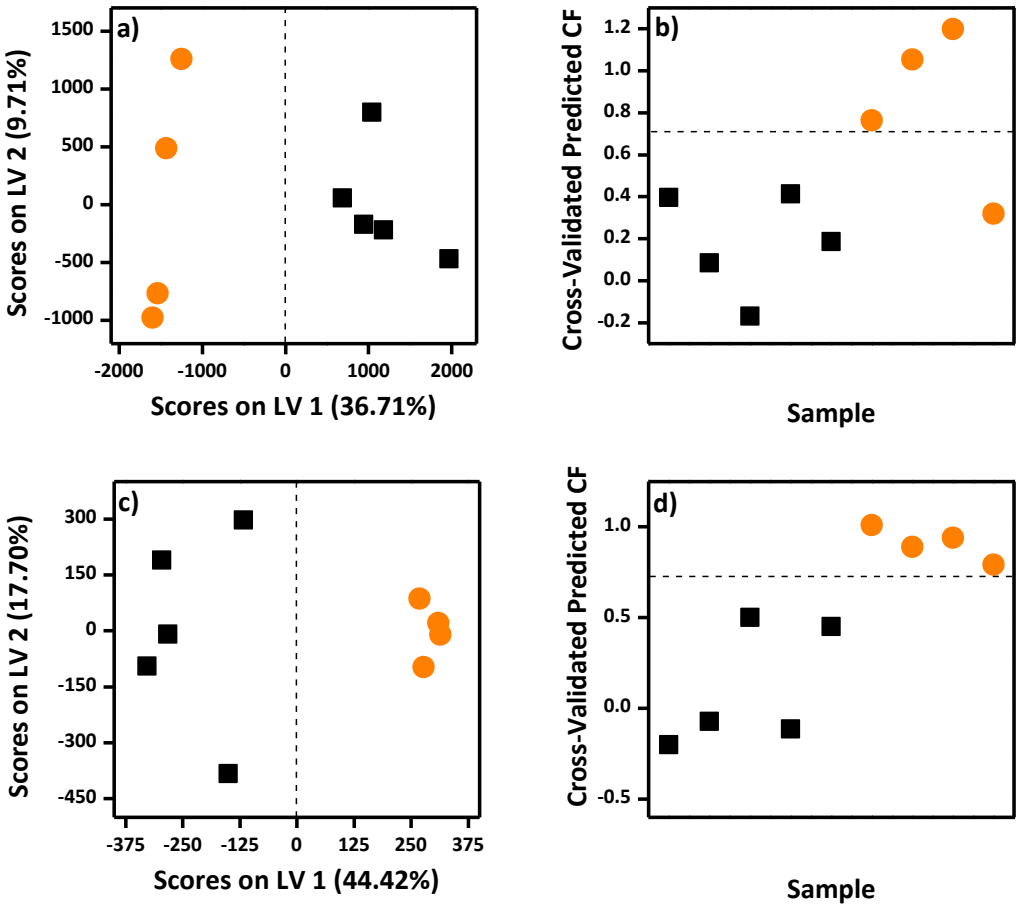

Figure 5.16: Orthogonal projection to latent structures-discriminant analysis (oPLS-DA) models for the discrimination of CF patients (orange circles) from healthy controls (black squares). (a): oPLS-DA calibration scores plot using the total initial set of 66 spectral features. The model consisted of 2 LVs with 46.43% and 93.79% total captured X- and Y-block variances, respectively. The cross-validated accuracy, sensitivity, and specificity were 88%, 100%, and 75%, respectively. (b): The corresponding CF cross-validated prediction plot for (a). There was 1 misclassified CF EBC sample. (c): oPLS-DA calibration scores plot using the 9 discriminant metabolic feature panel obtained from genetic algorithm variable selection. The model consisted of 2 LVs with 84.24% and 94.29% total captured X- and Y-block variances, respectively. The accuracy, sensitivity, and specificity were all 100%. (d): The corresponding CF cross-validated prediction plot for (c). There were no misclassified samples.

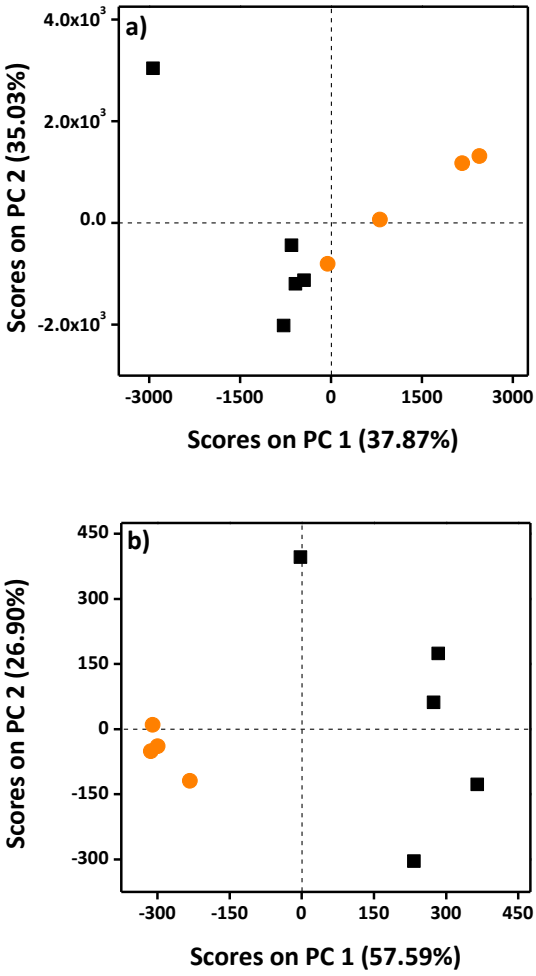

Figure 5.17: Principal Component Analysis (PCA) of cystic fibrosis (CF) patients (orange circles) and healthy controls (black squares). (a): PCA scores plot using the initial set of 66 spectral features. The model consisted of 2 PCs with 72.89% total captured variance. (b): PCA scores plot using the 9 discriminant feature panel obtained from genetic algorithm variable selection. The model consisted of 2 PCs with 84.49% total captured variance.

To further investigate these results, PCA was utilized to evaluate the performance of the set of 9 discriminant features in an unsupervised manner. Scores plots were generated for both the initial set of 66 metabolic features and the set of 9 discriminant features (Figure 5.17). Using the initial set, 2 principal components interpreting 72.89% of the data matrix variance showed some clustering of sample classes, but each class contained outliers (Figure 5.17a). However, better clustering of CF patients was observed with the 9-feature 2-principal component PCA model that interpreted 84.49 % of the variance (Figure 5.17b), providing further evidence of the robustness of the set of 9 discriminant features. TM-DART-TWIMS-TOF MS was able to successfully detect CF in this small sample cohort.

Although not all of the discriminant metabolites tentatively identified (Table 5.5) have been reported to be present in EBC, two have been linked to research focusing on CF. Grasemann et al. found that asymmetric dimethylarginine is increased in the airways of CF patients and may contribute to airway obstruction by reducing nitric oxide formation.[42] The authors suggested the endogenous hormone melatonin could aid in the reduction of asymmetric dimethylarginine in CF airways as *in vivo* experiments in rats have shown that melatonin treatment can decrease concentrations of asymmetric dimethylarginine in the liver, kidney, and blood plasma and increase nitric oxide formation.[43-45] Higher levels of N-acetylserotonin sulfate, a melatonin metabolite, were observed in our cohort of CF patients (Table 5.5), suggesting that melatonin was more rapidly metabolized by the CF patients. Bacteria species belonging to the genus *Inquilinus* have been isolated from CF patients. Chiron et al. found that one isolated strain was able to use valeric acid as its sole carbon source during assimilation testing.[46]

2-Hydroxyvaleric acid, a valeric acid oxidation product, was detected as a CF discriminatory metabolite in this study (Table 5.5). Many discriminant metabolites were flavonoids and other dietary metabolites with endogenous functions. Flavonoids have been shown to increase the chloride conductance of CFTR with the ΔF508 mutation (which is found in 70% of CF patients).[40]

Table 5.5: Tentative Metabolites Identified as Discriminatory Between Cystic Fibrosis Patients and Healthy Controls.

Feature Code	Drift time (ms)	m/z	Fold Change[1]	*P* value[2]	Ion type	Elemental Formula	Theoretical m/z	Δm (mDa)	Tentative Metabolite Identification
210	2.62	219.1391	--	0.39	$[M-H]^-$	$C_{14}H_{20}O_2$	219.1390	0.046	(Z)-6-Tetradecene-1,3-diyne-5,8-diol 13-Tetradecene-1,3-diyne-6,7-diol
217	2.68	184.0532	-2.96	0.043	$[M-H]^-$	-	-	-	N/A
240	3.28	243.1788	-1.93	0.08	$[M-H]^-$	$C_{17}H_{24}O$	243.1754	3.4	Panaxynol
						$C_{12}H_{24}N_2O_3$	243.1714	7.4	Isoleucyl-Leucine (& isomers)
251	2.07	118.0531	-1.39	0.01	$[M-H]^-$	$C_4H_9NO_3$	118.0509	2.1	Threonine Homoserine 2-Methyl-Serine
273	3.36	297.0575	0.83	0.26	$[M-H]^-$	$C_{12}H_{14}N_2O_5S$	297.0550	2.4	N-Acetylserotonin sulfate
						$C_{13}H_{14}O_8$	297.0615	4.1	Benzoyl glucuronide (Benzoic acid)
300	4.45	421.1654	6.70	0.049	$[M-H]^-$	$C_{25}H_{26}O_6$	421.1656	0.26	Mulberrin Kuwanon F Glyurallin B Cyclomammeisin Kuwanon D Mammea A/AB cyclo F

Table 5.5 (continued).

328	2.30	117.0452	0.77	0.12	$[M-H]^-$	$C_5H_{10}O_3$	117.0557	10.5	3-Hydroxy-2-methyl-[S-(R,R)]-butanoic acid 3-Hydroxy-2-methyl-[R-(R,S)]-butanoic acid 2-Hydroxyvaleric acid 3-Hydroxyvaleric acid 3-Hydroxyisovaleric acid 2-Hydroxy-2-methylbutyric acid 2-Hydroxy-3-methylbutyric acid Erythronilic acid 2-Methyl-3-hydroxybutyric acid Methyl 3-hydroxybutyrate
						C_8H_6O	117.0345	10.6	3,5,7-Octatriyn-1-ol 2,4,6-Octatriyn-1-ol xi-2,3-Octadiene-5,7-diyn-1-ol
413	3.19	252.1718	-4.11	0.030	$[M-H]^-$	--	--	--	N/A

Table 5.5 (continued).

436	2.44	125.0984	-4.00	0.025	$[M-H]^-$	$C_8H_{14}O$	125.0971	1.2	(3xi,5Z)-1,5-Octadien-3-ol 2,5-Octadien-1-ol (2E,4E)-2,4-Octadien-1-ol

[1]Fold change was calculated as the base 2 logarithm of the average peak area ratios for CF samples and control samples. Entries missing values are due to the inability to divide by zero. [2]*P* values calculated using Welch's *t*-test

5.8 Conclusion

TM-DART MS is a powerful analytical technique for metabolic fingerprinting. With optimal desorption temperatures and gas flow rates for efficient ionization and ion transmission, this approach detects a broad mass range of species. Furthermore, TM-DART MS has the ability to acquiring fragmentation data along with full scan MS data in a single experiment which minimizes sample consumption and allows for improved elemental formula assignment confidence. TM-DART coupled to TWIMS-TOF MS successfully detected CF in a small sample cohort, thereby, demonstrating it can be employed for probing metabolome changes. Comparing data acquired by TM-DART MS to that of the more commonly used chromatographic-MS platforms will provide a better perspective on the relevance and amount of biochemical information derived from TM-DART MS experiments.

5.9 References

(1) Cody, R. B.; Laramée, J. A.; Durst, H. D., Versatile new ion source for the analysis of materials in open air under ambient conditions. *Anal. Chem.* **2005,** *77* (8), 2297-2302.

(2) Harris, G. A.; Galhena, A. S.; Fernández, F. M., Ambient sampling/ionization mass spectrometry: applications and current trends. *Anal. Chem.* **2011,** *83* (12), 4508-4538.

(3) Song, L.; Gibson, S. C.; Bhandari, D.; Cook, K. D.; Bartmess, J. E., Ionization mechanism of positive-ion direct analysis in real time: a transient microenvironment concept. *Anal. Chem.* **2009,** *81* (24), 10080-10088.

(4) Zhou, M.; Guan, W.; Walker, L. D.; Mezencev, R.; Benigno, B. B.; Gray, A.; Fernandez, F. M.; McDonald, J. F., Rapid mass spectrometric metabolic profiling of blood sera detects ovarian cancer with high accuracy. *Cancer Epidemiol. Biomarkers Prev.* **2010,** *19* (9), 2262-2271.

(5) Cajka, T.; Riddellova, K.; Tomaniova, M.; Hajslova, J., Ambient mass spectrometry employing a DART ion source for metabolomic fingerprinting/profiling: a powerful tool for beer origin recognition. *Metabolomics* **2011,** *7* (4), 500-508.

(6) Dove, A. D.; Leisen, J.; Zhou, M.; Byrne, J. J.; Lim-Hing, K.; Webb, H. D.; Gelbaum, L.; Viant, M. R.; Kubanek, J.; Fernández, F. M., Biomarkers of whale shark health: a metabolomic approach. *PLoS One* **2012,** *7* (11), e49379.

(7) Gu, H.; Pan, Z.; Xi, B.; Asiago, V.; Musselman, B.; Raftery, D., Principal component directed partial least squares analysis for combining nuclear magnetic resonance and mass spectrometry data in metabolomics: application to the detection of breast cancer. *Anal. Chim. Acta* **2011,** *686* (1-2), 57-63.

(8) Kim, S. W.; Kim, H. J.; Kim, J. H.; Kwon, Y. K.; Ahn, M. S.; Jang, Y. P.; Liu, J. R., A rapid, simple method for the genetic discrimination of intact Arabidopsis thaliana mutant seeds using metabolic profiling by direct analysis in real-time mass spectrometry. *Plant Methods* **2011,** *7*, 14.

(9) Harris, G. A.; Fernández, F. M., Simulations and experimental investigation of atmospheric transport in an ambient metastable-induced chemical ionization source. *Anal. Chem.* **2008,** *81* (1), 322-329.

(10) Perez, J. J.; Harris, G. A.; Chipuk, J. E.; Brodbelt, J. S.; Green, M. D.; Hampton, C. Y.; Fernandez, F. M., Transmission-mode direct analysis in real time and desorption

electrospray ionization mass spectrometry of insecticide-treated bednets for malaria control. *Analyst* **2010,** *135* (4), 712-719.

(11) Zhou, M.; McDonald, J. F.; Fernandez, F. M., Optimization of a direct analysis in real time/time-of-flight mass spectrometry method for rapid serum metabolomic fingerprinting. *J. Am. Soc. Mass. Spectrom.* **2010,** *21* (1), 68-75.

(12) Dettmer, K.; Aronov, P. A.; Hammock, B. D., Mass spectrometry-based metabolomics. *Mass Spectrom. Rev.* **2007,** *26* (1), 51-78.

(13) Lenz, E. M.; Wilson, I. D., Analytical strategies in metabonomics. *J. Proteome Res.* **2007,** *6* (2), 443-458.

(14) Wishart, D. S., Quantitative metabolomics using NMR. *Trac-Trend Anal Chem* **2008,** *27* (3), 228-237.

(15) Nyadong, L.; Galhena, A. S.; Fernandez, F. M., Desorption electrospray/metastable-induced ionization: a flexible multimode ambient ion generation technique. *Anal. Chem.* **2009,** *81* (18), 7788-7794.

(16) Davis, J. M.; Calvin Giddings, J., Statistical theory of component overlap in multicomponent chromatograms. *Anal. Chem.* **1983,** *55* (3), 418-424.

(17) Creaser, C. S.; Griffiths, J. R.; Bramwell, C. J.; Noreen, S.; Hill, C. A.; Thomas, C. L. P., Ion mobility spectrometry: a review. Part 1. Structural analysis by mobility measurement. *Analyst* **2004,** *129* (11), 984-994.

(18) Mason, E. A.; Schamp, H. W., Mobility of gaseous lons in weak electric fields. *Ann. Phys. New York* **1958,** *4* (3), 233-270.

(19) Kaplan, K.; Dwivedi, P.; Davidson, S.; Yang, Q.; Tso, P.; Siems, W.; Hill Jr, H. H., Monitoring dynamic changes in lymph metabolome of fasting and fed rats by electrospray ionization-ion mobility mass spectrometry (ESI-IMMS). *Anal. Chem.* **2009,** *81* (19), 7944-7953.

(20) Dwivedi, P.; Schultz, A. J.; Hill Jr, H. H., Metabolic profiling of human blood by high-resolution ion mobility mass spectrometry (IM-MS). *Int. J. Mass spectrom.* **2010,** *298* (1-3), 78-90.

(21) Kaplan, K. A.; Chiu, V. M.; Lukus, P. A.; Zhang, X.; Siems, W. F.; Schenk, J. O.; Hill Jr, H. H., Neuronal metabolomics by ion mobility mass spectrometry: cocaine effects on glucose and selected biogenic amine metabolites in the frontal cortex, striatum, and thalamus of the rat. *Anal. Bioanal. Chem.* **2013,** *405* (6), 1959-1968.

(22) McFarland, M.; Albritton, D.; Fehsenfeld, F.; Ferguson, E.; Schmeltekopf, A. L., Flow-drift technique for ion mobility and ion-molecule reaction rate constant

measurements. I. Apparatus and mobility measurements. *J. Chem. Phys.* **1973,** *59* (12), 6610-6619.

(23) Buryakov, I.; Krylov, E.; Nazarov, E.; Rasulev, U. K., A new method of separation of multi-atomic ions by mobility at atmospheric pressure using a high-frequency amplitude-asymmetric strong electric field. *Int. J. Mass Spectrom. Ion Processes* **1993,** *128* (3), 143-148.

(24) Giles, K.; Pringle, S. D.; Worthington, K. R.; Little, D.; Wildgoose, J. L.; Bateman, R. H., Applications of a travelling wave-based radio-frequency-only stacked ring ion guide. *Rapid Commun. Mass Spectrom.* **2004,** *18* (20), 2401-2414.

(25) Lanucara, F.; Holman, S. W.; Gray, C. J.; Eyers, C. E., The power of ion mobility-mass spectrometry for structural characterization and the study of conformational dynamics. *Nat. Chem.* **2014,** *6* (4), 281-294.

(26) Wishart, D. S.; Jewison, T.; Guo, A. C.; Wilson, M.; Knox, C.; Liu, Y.; Djoumbou, Y.; Mandal, R.; Aziat, F.; Dong, E., HMDB 3.0—the human metabolome database in 2013. *Nucleic Acids Res.* **2012,** *41* (Supp 1), D801-D807.

(27) Pluskal, T.; Castillo, S.; Villar-Briones, A.; Orešič, M., MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* **2010,** *11*, 395.

(28) Zhou, B.; Wang, J.; Ressom, H. W., MetaboSearch: tool for mass-based metabolite identification using multiple databases. *PLoS One* **2012,** *7* (6), e40096.

(29) Smith, C. A.; O'Maille, G.; Want, E. J.; Qin, C.; Trauger, S. A.; Brandon, T. R.; Custodio, D. E.; Abagyan, R.; Siuzdak, G., METLIN: a metabolite mass spectral database. *Ther. Drug Monit.* **2005,** *27* (6), 747-751.

(30) Fahy, E.; Subramaniam, S.; Murphy, R. C.; Nishijima, M.; Raetz, C. R.; Shimizu, T.; Spener, F.; van Meer, G.; Wakelam, M. J.; Dennis, E. A., Update of the LIPID MAPS comprehensive classification system for lipids. *J. Lipid Res.* **2009,** *50* (Supplement), S9-S14.

(31) Harris, G. A.; Hostetler, D. M.; Hampton, C. Y.; Fernandez, F. M., Comparison of the internal energy deposition of direct analysis in real time and electrospray ionization time-of-flight mass spectrometry. *J. Am. Soc. Mass. Spectrom.* **2010,** *21* (5), 855-63.

(32) Pierce, C. Y.; Barr, J. R.; Cody, R. B.; Massung, R. F.; Woolfitt, A. R.; Moura, H.; Thompson, H. A.; Fernandez, F. M., Ambient generation of fatty acid methyl ester ions from bacterial whole cells by direct analysis in real time (DART) mass spectrometry. *Chem. Commun.* **2007,** (8), 807-809.

(33) Nilles, J. M.; Connell, T. R.; Durst, H. D., Thermal separation to facilitate Direct Analysis in Real Time (DART) of mixtures. *Analyst* **2010,** *135* (5), 883-886.

(34) Biggs, W. D.; McColl, I. R.; Moon, J. R., Electrical and Thermal Conductivity. In *Construction Materials: Their Nature and Behaviour*, 3rd ed.; Illston, J. M.; Domone, P. L. J., Eds. Taylor & Francis: London, 2001.

(35) Database, N. C. M. P., http://www.cryogenics.nist.gov/. In.

(36) Sumner, L.; Amberg, A.; Barrett, D.; Beale, M.; Beger, R.; Daykin, C.; Fan, T.; Fiehn, O.; Goodacre, R.; Griffin, J.; Hankemeier, T.; Hardy, N.; Harnly, J.; Higashi, R.; Kopka, J.; Lane, A.; Lindon, J.; Marriott, P.; Nicholls, A.; Reily, M.; Thaden, J.; Viant, M., Proposed minimum reporting standards for chemical analysis. *Metabolomics* **2007,** *3* (3), 211-221.

(37) Moco, S.; Vervoort, J.; Moco, S.; Bino, R. J.; De Vos, R. C. H.; Bino, R., Metabolomics technologies and metabolite identification. *TrAC, Trends Anal. Chem.* **2007,** *26* (9), 855-866.

(38) Olivera, A.; Spiegel, S., Sphingosine-1-phosphate as second messenger in cell proliferation induced by PDGF and FCS mitogens. *Nature* **1993,** *365* (6446), 557-560.

(39) Pyne, S.; Pyne, N. J., Sphingosine 1-phosphate signaling in mammalian cells. *Biochem. J* **2000,** *349*, 385-402.

(40) Lim, M.; McKenzie, K.; Floyd, A. D.; Kwon, E.; Zeitlin, P. L., Modulation of ΔF508 cystic fibrosis transmembrane regulator trafficking and function with 4-phenylbutyrate and flavonoids. *Am. J. Respir. Cell Mol. Biol.* **2004,** *31* (3), 351-357.

(41) Wetmore, D. R.; Joseloff, E.; Pilewski, J.; Lee, D. P.; Lawton, K. A.; Mitchell, M. W.; Milburn, M. V.; Ryals, J. A.; Guo, L., Metabolomic profiling reveals biochemical pathways and biomarkers associated with pathogenesis in cystic fibrosis cells. *J. Biol. Chem.* **2010,** *285* (40), 30516-30522.

(42) Grasemann, H.; Al-Saleh, S.; Scott, J. A.; Shehnaz, D.; Mehl, A.; Amin, R.; Rafii, M.; Pencharz, P.; Belik, J.; Ratjen, F., Asymmetric dimethylarginine contributes to airway nitric oxide deficiency in patients with cystic fibrosis. *Am. J. Respir. Crit. Care Med.* **2011,** *183* (10), 1363-1368.

(43) Tain, Y.-L.; Kao, Y.-H.; Hsieh, C.-S.; Chen, C.-C.; Sheen, J.-M.; Lin, I.-C.; Huang, L.-T., Melatonin blocks oxidative stress-induced increased asymmetric dimethylarginine. *Free Radical Biol. Med.* **2010,** *49* (6), 1088-1098.

(44) Tain, Y. L.; Hsieh, C. S.; Chen, C. C.; Sheen, J. M.; Lee, C. T.; Huang, L. T., Melatonin prevents increased asymmetric dimethylarginine in young rats with bile duct ligation. *J. Pineal Res.* **2010,** *48* (3), 212-221.

(45) Tain, Y. L.; Huang, L. T.; Lin, I.; Lau, Y. T.; Lin, C. Y., Melatonin prevents hypertension and increased asymmetric dimethylarginine in young spontaneous hypertensive rats. *J. Pineal Res.* **2010,** *49* (4), 390-398.

(46) Chiron, R.; Marchandin, H.; Counil, F.; Jumas-Bilak, E.; Freydiere, A.-M.; Bellon, G.; Husson, M.-O.; Turck, D.; Brémont, F.; Chabanon, G., Clinical and microbiological features of Inquilinus sp. isolates from five patients with cystic fibrosis. *J. Clin. Microbiol.* **2005,** *43* (8), 3938-3943.

CHAPTER 6. CONCLUSIONS AND OUTLOOK

6.1 Abstract

This chapter presents a summary of the results pertaining to the use of mass spectrometry (MS)-based metabolomics for cancer detection and the study of ecological interactions in addition to the development of TM-DART for metabolomic fingerprinting discussed in this dissertation. Additionally, potential future directions for continuation of this work are given.

6.2 Ultra Performance Liquid Chromatography-Mass Spectrometry Serum Metabolomics for Cancer Detection

6.2.1 Summary of Accomplishments

The major finding in the work presented in this dissertation is the feasibility of MS-based untargeted metabolomics for prostate (PCa) and ovarian cancer (OC) detection. A metabolite-based *in vitro* diagnostic multivariate index assay (IVDMIA) created from UPLC-MS/MS serum metabolomic analysis of 64 PCa patients and 50 healthy individuals coupled to machine learning methods predicted the presence of PCa with high classification sensitivity, specificity and accuracy. A panel of 40 metabolic features was found to be differential with 92.1% sensitivity, 94.3% specificity, and 93.0% accuracy. Of further significance, the IVDMIA was able to correctly predict 100% of the true positives that were incorrectly diagnosed as negatives by the conventional PSA test; highlighting that a combination of multiple discriminant features yields higher predictive power for PCa detection than the univariate analysis of a single biomarker. Within the

discriminant panel, 31 metabolites were identified by high-resolution accurate mass MS and MS/MS, with 10 further confirmed chromatographically by standards. Fatty acids, amino acids, lysophospholipids, and bile acids have been identified among the discriminant metabolites, suggesting alterations in their metabolism. Additionally, several metabolites were mapped to the steroid hormone biosynthesis pathway. These observations demonstrate some of the plausible metabolic alterations in PCa, and provide further insight into the biological pathway changes associated with the disease. If higher throughput analysis, and lower analysis cost and complexity are desired, 13 metabolites that were found to be present in 90% of the entire sample cohort would still provide high classification sensitivity (88.3%), specificity (80.3%), and accuracy (85.0%) for cancerous and healthy samples.

Two systems were used to investigate metabolic patterns for early detection of OC. In the first system, early-stage HGSC was successfully detected in a *Dicer-Pten* double knockout (DKO) mouse model utilizing UPLC-MS untargeted metabolic fingerprinting. After down-selection of metabolic features with maximum discriminatory power, 18 metabolites differentiated 14 DKO mice with early-stage tumors from 11 control mice with 100% accuracy, sensitivity and specificity. Altered metabolic pathways included those of fatty acids, bile acids and alcohols, glycerophospholipids, peptides, and phytochemicals. These alterations impact cellular energy storage and membrane stability, as well as defenses against oxidative stress. Building upon previous work, the second system studied showed that early-stage OC detection in humans employing UPLC-MS untargeted metabolic fingerprinting is feasible. After down-selection of metabolic features with maximum discriminatory power by recursive feature elimination, 22

metabolites differentiated 24 early-stage OC patients from 40 healthy women with 100% cross-validated accuracy, sensitivity, and specificity. Although this was a small pilot study—poor early diagnosis currently complicates collection of large patient cohorts for more detailed studies—it demonstrates that metabolic information can be useful for detecting early-stage OC.

6.2.2 Moving Forward

While the presented PCa and OC metabolic fingerprinting studies demonstrate the applicability of MS-based untargeted metabolomics for cancer detection, additional studies are needed to further validate and test the robustness of the proposed metabolite-based detection panels. The following studies are recommended for future research:

1. A drawback of the PCa and OC studies is the number of patients involved. Larger-scale studies are needed to prove that these metabolites are indeed indicators of these cancers across a broad population. According to formulas proposed by Arkin and Wachtel, a biomarker discovery study that aims to have a fixed specificity of 95% and sensitivity of at least 95% (±0.05) at the 95% confidence interval should include 73 diseased patients at minimum.[1] Xia et al. advises matching each diseased patient to four random healthy controls for low-prevalence diseases such as OC which affects 1.3% of women.[2, 3] The authors also suggest multiplying the minimum sample size by 1.5 if multivariate models will be cross-validated. Therefore, using the OC metabolomics detection work as an example, at least 110 early-stage OC patients and 440 controls (healthy individuals) are needed for biomarker discovery studies. Currently, David Gaul (a research scientist in the Fernández research group) is pursuing an additional

metabolomics study for detection of early-stage OC in a larger patient cohort that builds upon what was demonstrated in the pilot human study presented in Chapter 3. Additionally, María Eugenia Monge (a prior research scientist in the Fernández research group) has engaged in a collaboration with Hospital Italiano de Buenos Aires to determine if the findings presented in Chapter 2 still hold when applied to a cohort of 500 PCa patients of Argentine descent.

2. The cancer detection studies presented were based on untargeted metabolomic fingerprinting methodology, but translation to a clinical setting will require the development of targeted assays. Tandem MS (as used for newborn screening of inherited metabolic disorders) provides the specificity and sensitivity needed for such targeted assays.[4] Selected reaction monitoring (SRM) or multiple reaction monitoring (MRM) can be utilized to detect only user-specified precursor metabolite ions and the corresponding fragment product ion(s) (referred to as transitions). Experiments must be conducted to detect and ascertain the transitions of the discriminant metabolites determined in these studies to facilitate their use in a clinical setting.
3. Only a fraction of the metabolic information that was amassed during these cancer detection studies was utilized. While primary aims were focused on demonstrating the efficacy of using metabolomics for biomarker discovery/cancer detection, the wealth of information garnered should also be utilized to investigate the biological processes and proliferative molecular mechanisms of these cancers. Instead of focusing on metabolite panels, all detected metabolites with significant concentration changes between groups should be mapped to further understand

the impact each cancer has metabolic pathways. A better understanding of the biological processes and proliferative molecular mechanisms can lead to more targets for chemotherapeutics. In this vein, this work also stands to benefit majorly from the integration of this data with other omics data such as proteomics.

4. Preliminary MS imaging experiments of DKO mice oviduct tissue sections performed my Martin Paine (a postdoctoral researcher in the Fernández research group) have shown that chemically imaging metabolite abundances may provide useful information regarding tumor formation-associated metabolite abundance changes. Conducting similar experiments with cancerous tissue samples from humans may help in understanding the underlying metabolic perturbations caused by tumor development in both OC and PCa.

For the OC *Dicer-Pten* DKO mouse model work specifically, an exploration of the shared metabolome overlap between DKO mice and humans at early stages of HGSC development is needed to further understand how relevant scientific findings resulting from research using this mouse model can carry over to the diseased human population. Moreover, as metabolomics is actively used to understand the molecular mechanisms by which chemotherapeutic drugs attack and destroy cancerous cells in addition to assessing the efficacy of treatment, the *Dicer-Pten* DKO mice can be used to gather metabolic fingerprints related to the efficacy of new chemotherapeutic drugs or to gain insights regarding the molecular mechanisms of their interaction with their targets since these mice provide a simpler, better-controlled model to study these effects.

6.3 Metabolic Impacts of Chemically Mediated Competition on Marine Plankton

6.3.1 Summary of Accomplishments

The work presented in Chapter 4 represents the first instance of metabolites and proteins measured simultaneously to understand the effects of allelopathy or any form of competition among diatoms. Whole cell metabolomic and proteomic analyses revealed differing responses of planktonic competitors to *Karenia brevis* allelopathy, suggesting that co-occurring species may have evolved partial resistance to allelopathy via robust metabolic pathways. In contrast, a "naïve" competitor, *Thalassiosira pseudonana*, which does not co-occur with *K. brevis* blooms, suffered greater metabolic disruption and growth suppression when exposed to *K. brevis* allelopathy. Critical metabolic processes including glycolysis, photosynthesis, cell membrane maintenance, osmoregulation, and responses to oxidative stress were all impacted in *T. pseudonana*. This interdisciplinary systems biology approach provided an unbiased opportunity to establish novel, testable hypotheses towards understanding the mechanisms by which allelopathy alters species composition in plankton communities, and, thus, the roles of chemical cues in mediating important ecological interactions.

6.3.2 Moving Forward

K. brevis blooms pose threats to human health and destroy endangered marine mammal populations at an estimated cost of more than $18 million per bloom.[5, 6] Knowledge of the mechanisms by which allelopathy alters species composition in plankton communities, and, thus, the roles of chemical cues in mediating important

ecological interactions potentially allow for a deeper understanding of bloom dynamics that may lead to prediction and mitigation strategies to alleviate their harmful effects. While this work exemplified how the integration of data from systems biology fields can characterize physiological responses to stimuli and generate hypotheses related to cellular function, more can be done to further understand how *K. brevis* allelopathy affects competitor species. The following studies are recommended for future research:

1. The presented work focused primarily on polar metabolites (i.e., non-lipids). Yet, plankton lipids mediate a number of important ecological interactions, including grazing,[7] susceptibility to viral infection,[8] and toxin sensitivity.[9] Thus, future experiments testing how phytoplankton competition alters lipid biochemistry are needed. These experiments could lead to discovery of cascading ecological effects on other types of interactions. Remy Poulin (a graduate student in the Kubanek research group) has begun these experiments using NMR metabolomics.
2. Thus far, the impact of *K. brevis* allelopathy on diatom metabolism was only investigated for one co-occuring competitor species—*Assterionellopsis glacialis*. Experiments should be conducted to understand the metabolic impacts on other co-occuring competitor species such as *Skeletonema grethae*, *Odontella aurita*, and *Stephanopyxis turris*.[10] These experiments can produce further findings that generate hypotheses towards understanding the mechanisms by which allelopathy alters species composition in plankton communities.

6.4 Transmission Mode Direct Analysis in Real Time for Metabolic Fingerprinting

6.4.1 Summary of Accomplishments

Although current hyphenated MS techniques are very powerful for metabolomics studies, there is a clear need for more rapid, high-throughput MS approaches. TM-DART MS was shown to be a powerful, rapid analytical technique for metabolic fingerprinting. With optimal desorption temperatures and gas flow rates for efficient ionization and ion transmission, this approach detects a broad mass range of metabolite ionic species. Ramping the ionizing gas desorption temperature further enhanced analysis by adding a simple separation dimension to this ambient approach. Furthermore, TM-DART MS has the ability to acquire fragmentation data along with full scan MS data in a single experiment which minimizes sample consumption and allows for improved elemental formula assignment confidence. In terms of reproducibility, monitored TM-DART signals from a quinine standard resulted in coefficients of variation as low as 16%. Moreover, TM-DART coupled to TWIMS-TOF MS successfully detected CF in a small sample cohort, thereby, demonstrating it can be employed for probing metabolome changes.

6.4.2 Moving Forward

While TM-DART was successfully used to probe metabolome changes, more experiments are needed to establish the technique as viable for large-scale metabolomics studies. The following studies are recommended for future research:

1. Similarly to GC-MS analysis, serum samples used in the TM-DART method development work were derivatized via silylation reactions. Chemical derivatization reactions can create undesirable artifacts and also have varying degrees of efficiency as sterically hindered analytes may only be partially derivatized.[11, 12] Moreover, this added sample preparation step decreases sample throughput, thus, increasing analysis time which may reduce the advantage of the inherent rapidness of the technique. As the EBC analyses did not require a derivatization sample preparation step, experiments should be conducted to evaluate the ability of TM-DART to analyze non-derivatized serum samples.
2. Many biological sample types are utilized for metabolomic fingerprinting studies. The ability of TM-DART MS to acquire discriminant metabolic information from a multitude of sample types should be evaluated. This can be used to further show the applicability of TM-DART for many different metabolomic fingerprinting studies.
3. Data acquired by TM-DART MS should be compared to that of the more commonly used chromatographic-MS platforms so that a better perspective on the relevance and amount of biochemical information derived from TM-DART MS experiments can be provided. José J. Pérez, a graduate student in the Fernández research group, is currently comparing the chemical information acquired from TM-DART MS analysis of EBC samples to that of direct infusion ESI and APCI MS. But, an evaluation of how TM-DART MS fares in comparison to LC-MS in answering a metabolomic question would be useful.

6.5 References

(1) Arkin, C. F.; Wachtel, M. S., How many patients are necessary to assess test performance? *J. Am. Med. Assoc.* **1990,** *263* (2), 275-278.

(2) Xia, J.; Broadhurst, D. I.; Wilson, M.; Wishart, D. S., Translational biomarker discovery in clinical metabolomics: an introductory tutorial. *Metabolomics* **2013,** *9* (2), 280-299.

(3) Siegel, R. L.; Miller, K. D.; Jemal, A., Cancer statistics, 2015. *CA Cancer J. Clin.* **2015,** *65* (1), 5-29.

(4) Li, W.; Tse, F. L., Dried blood spot sampling in combination with LC-MS/MS for quantitative analysis of small molecules. *Biomed. Chromatogr.* **2010,** *24* (1), 49-65.

(5) Anderson, D. M.; Garrison, D. J., *The ecology and oceanography of harmful algal blooms*. American Society of Limnology and Oceanography: Waco, TX, 1997.

(6) Landsberg, J. H., The effects of harmful algal blooms on aquatic organisms. *Rev. Fish. Sci.* **2002,** *10* (2), 113-390.

(7) Miralto, A.; Barone, G.; Romano, G.; Poulet, S. A.; Ianora, A.; Russo, G. L.; Buttino, I.; Mazzarella, G.; Laabir, M.; Cabrini, M.; Giacobbe, M. G., The insidious effect of diatoms on copepod reproduction. *Nature* **1999,** *402* (6758), 173-176.

(8) Vardi, A.; Van Mooy, B. A. S.; Fredricks, H. F.; Popendorf, K. J.; Ossolinski, J. E.; Haramaty, L.; Bidle, K. D., Viral glycosphingolipids induce lytic infection and cell death in marine phytoplankton. *Science* **2009,** *326* (5954), 861-865.

(9) Deeds, J. R.; Place, A. R., Sterol-specific membrane interactions with the toxins from *Karlodinium micrum* (Dinophyceae) — a strategy for self-protection? *Afr. J. Mar. Sci.* **2006,** *28* (2), 421-425.

(10) Poulson-Ellestad, K.; Mcmillan, E.; Montoya, J. P.; Kubanek, J., Are offshore phytoplankton susceptible to Karenia brevis allelopathy? *J. Plankton Res.* **2014,** *36* (5), 1344-1356.

(11) Halket, J. M.; Waterman, D.; Przyborowska, A. M.; Patel, R. K.; Fraser, P. D.; Bramley, P. M., Chemical derivatization and mass spectral libraries in metabolic profiling by GC/MS and LC/MS/MS. *J. Exp. Bot.* **2005,** *56* (410), 219-243.

(12) Little, J. L., Artifacts in trimethylsilyl derivatization reactions and ways to avoid them. *J. Chromatogr. A* **1999,** *844* (1), 1-22.

VITA

CHRISTINA M. JONES

CHRISTINA M. JONES was born and raised in Baton Rouge, Louisiana. She attended public schools in East Baton Rouge Parish and graduated from Baton Rouge Magnet High School in May 2005. In the summer of 2005 she enrolled at Louisiana State University and A&M College in Baton Rouge. She graduated in May 2009 with a B.S. in Chemistry and minors in Sociology and African & African American Studies. In August 2009 she moved to Atlanta, Georgia to pursue a doctorate in Analytical Chemistry from the Georgia Institute of Technology. When she is not working on her research, Christina enjoys spending time with friends, listening to music, watching movies, reading, eating out, and exploring Atlanta.